Special report: Ukraine's 'hero city' of science p. 1036

The future of psychedelics in medicine p. 1051

Evolution of a native agricultural weed pp. 1053 & 1079

Science

$15
9 DECEMBER 2022
science.org

AAAS



CODE BY AI

Matching humans in programming competitions

pp. 1056 & 1092

VISIT…

LANZAROTE Caliente.com

YOUR RESEARCH HAS ITS REWARDS.



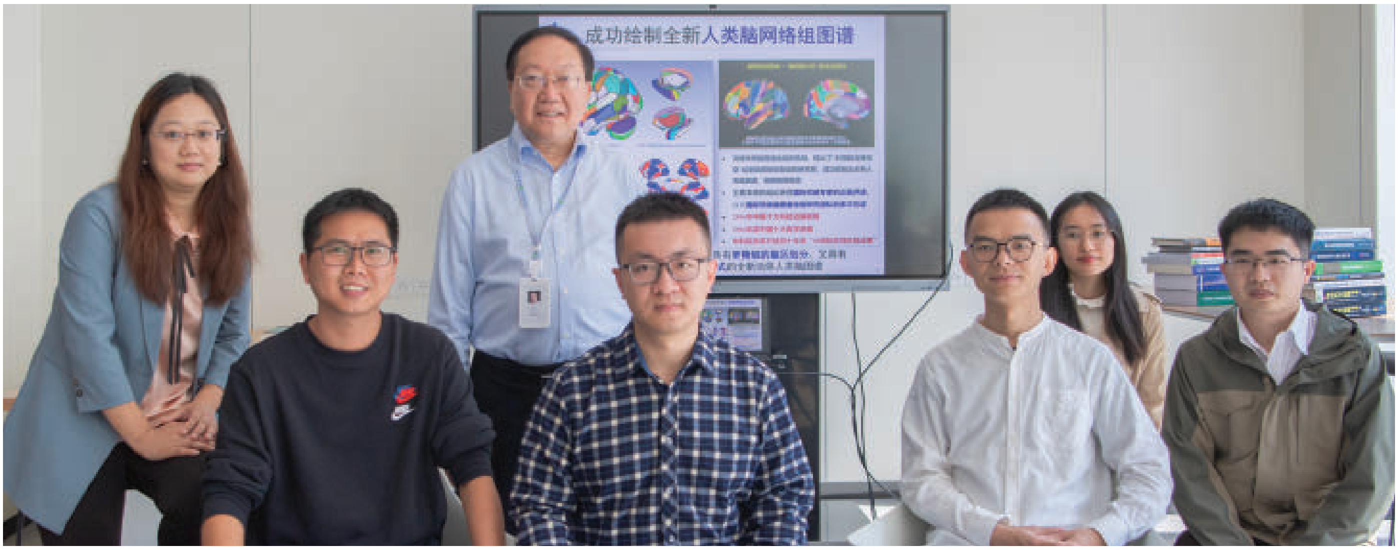
Submit an essay by 3/15/23 to enter to win $25K.

The prestigious international *Science* & PINS Prize is awarded annually to outstanding early-career researchers in the emerging field of neuromodulation. Findings already show improved outcomes in several neurological disorders. To enter, just write a 1000-word essay about your research performed in the last three years. The Grand Prize winner will be presented a US$25,000 award, as well as have their essay published in *Science* magazine and on *Science* Online. A runner-up will have their essay published online. For additional inquiries, email SciencePINSPrize@aaas.org.

SCIENCE-PINS PRIZE FOR NEUROMODULATION

SCIENCE-PINS

SCIENCE.ORG/PINS



Tianzi Jiang (back row, center) with the Zhejiang Lab Brain Intelligence Team

A bottom-up approach to brain-inspired AI

Scientists at Zhejiang Lab are learning more about neurobiology, and turning that knowledge into more powerful artificial intelligence.

In the 1940s, scientists started developing simple electrical circuits intended to mimic the interactions of neurons in the brain. Those early neural networks evolved into computer models of artificial intelligence (AI) in the 1950s. Since then, neuroscience and AI have progressed along parallel paths.

Tianzi Jiang is a senior research specialist at China-based Zhejiang Lab's Research Center for Human-Machine Augmented Intelligence; he and his colleagues are developing new brain-inspired AI systems driven by data and knowledge. By working on computability, the generalization of AI, and other areas, these scientists aim to improve AI cognition, decision making, and learning. For example, Zhejiang Lab has created the Large-Scale Humanoid Intelligence Knowledge Base, which can be applied to AI-based knowledge mining, memorization, and more.

"Recent scientific efforts have offered some insights by examining how the objectively measured patterns of brain organization and function can inspire the design of the artificial neural network," says Jiang.

To help scientists develop and test new methods and technologies, Zhejiang Lab recently built its Dubhe AI open-source platform. This system arises from a high-performance, distributed-computing architecture that integrates one-stop AI model development, a high-performance deep-learning framework, and a knowledge-refining platform. The features of this platform include a complete parallel mode, automatic arrangement and execution of computing steps, and high operational efficiency and stability. Scientists can apply this platform to research on AI-based learning and decision making, which drives a new-generation of AI technology based on human-like knowledge and autonomous learning.

"Instead of treating the AI system as a black box," says Jiang, "scientists at the Zhejiang Lab keep bottom-up options open to establish a biologically plausible AI system inspired by the multiscale principles of brain organization and function to gain access to the realm of biological intelligence."

A vision of new brain-inspired AI models

By studying the organization and activity patterns in the brain, scientists can incorporate that knowledge in new artificial neural networks. "A promising example comes from the correspondence between the activities of a convolutional neural network and the human ventral visual pathway during the processing of the same images," Jiang explains. "When the input images are specified to contain face stimuli only, the activity of the face-selective area in the human brain can be well predicted by the activity of face-selective units in the model."

Although better understanding of the brain promises even better AI-based applications, Jiang points out that "the brain is highly heterogeneous at different spatiotemporal scales, and exploring it requires an efficient approach—just like a nautical chart [is necessary] for a navigator—to integrate different pieces of knowledge to form a more complete landscape."

To provide such a chart, Jiang's team created the Brainnetome Atlas, which is described as "an in vivo map, with more fine-grained functional brain subregions and detailed anatomical and functional connection patterns for each area." From this, Jiang and his colleagues developed a Brainnetome-inspired AI system, in which he says a digital-twin brain can "achieve a biologically plausible neural network to understand and process both data and knowledge in a humanoid way."

Zhejiang Lab scientists are far ahead of their predecssors who built the neural networks of the 1940s—yet they are only getting started in advancing and applying brain-mimicking AI. As Jiang says, "By leveraging the knowledge accumulated in brain studies, more brain processes can be included to inspire the design of artificial neural networks, which will hopefully push the boundary of our understanding of human intelligence."

Sponsored by

之江实验室
ZHEJIANG LAB

PHOTO: PROVIDED BY ZHEJIANG LAB

CONTENTS



1036
Russian missiles and artillery pummeled Kharkiv, Ukraine's major science center, earlier this year.

9 DECEMBER 2022 • VOLUME 378 • ISSUE 6624

NEWS

IN BRIEF

1028 News at a glance

IN DEPTH

1030 Alzheimer's drug stirs excitement—and concerns
Antibody slows cognitive decline, but deaths, brain bleeds, and swelling mar results *By J. Couzin-Frankel and C. Piller*

1032 NASA radar altimetry mission to study hidden ocean swirls
Enhanced resolution of SWOT satellite will highlight how small eddies soak up heat and carbon *By P. Voosen*

1033 Image problems besiege Stanford president
Marc Tessier-Lavigne's early papers are subject of school and journal investigations *By J. Kaiser*

1034 In China, 'zero COVID' has become a Catch-22
Population chafes at control measures, but lifting them now would carry huge risks *By D. Normile*

1035 New U.S. law aims to light up medical research on cannabis
Biden signs bill streamlining pot studies and production *By M. Wadman*

FEATURES

1036 Hero city
Kharkiv was Ukraine's science hotbed until Russia attacked. The crippled city refuses to give up *By R. Stone*

1039 The volunteer *By R. Stone*

1040 The caretaker *By R. Stone*

PODCAST

1042 Out of the ashes
Slavutych was built after the nuclear meltdown at Chornobyl. The town, scarred by war, is planning a research revival *By R. Stone*

PODCAST

INSIGHTS

BOOKS *ET AL.*

1044 New books for young scientists

PERSPECTIVES

1051 The therapeutic potential of psychedelics
The development of psychedelics as medicines faces several challenges *By E. A. D. Schindler and D. C. D'Souza*

1053 Herbaria reveal cost of the green revolution
Rapid weed evolution is exposed by genome sequencing of natural history collections *By K. Waselkov and K. M. Olsen*

RESEARCH ARTICLE p. 1079

1054 Getting bifunctional molecules into cells
A class of transmembrane proteins helps shuttle large drugs across the cell membrane *By R. S. Lokey and C. Pye*

REPORT p. 1097



1044

CREDITS: (PHOTO) KYODO NEWS VIA AP IMAGES; (ILLUSTRATION) CHIARA VERCESI

1056 AlphaCode and "data-driven" programming
Is ignoring everything that is known about code the best way to write programs?
By J. Z. Kolter
RESEARCH ARTICLE p. 1092

POLICY FORUM

1057 Fairness considerations in global mitigation investments
Current mitigation finance flows are inadequate and unfair *By S. Pachauri* et al.

LETTERS

1060 Data-based tools can prevent crowd crushes
By M. Haghani and R. Lovreglio

1060 Co-management of Chile's escaped farmed salmon
By G. Figueroa-Muñoz et al.

1061 Long-term solutions for China's heat and drought
By X. Guo et al.

RESEARCH

IN BRIEF

1062 From *Science* and other journals

REVIEW

1065 Batteries
Designing better electrolytes *Y. S. Meng* et al.
REVIEW SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABQ3750

RESEARCH ARTICLES

1066 Systems biology
Principles of gene regulation quantitatively connect DNA to RNA and proteins in bacteria
R. Balakrishnan et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABK2066

1067 Computer science
Human-level play in the game of *Diplomacy* by combining language models with strategic reasoning *Meta Fundamental AI Research Diplomacy Team (FAIR)*

1074 Geology
A new biologic paleoaltimetry indicating Late Miocene rapid uplift of northern Tibet Plateau
Y. Miao et al.



Common waterhemp (*Amaranthus tuberculatus*) is native to North America, where it has become a prominent agricultural weed in the last half century as it evolved in response to intensive farming practices.

1079 Evolutionary ecology
Rapid weed adaptation and range expansion in response to agriculture over the past two centuries *J. M. Kreiner* et al.
PERSPECTIVE p. 1053

1085 Organic chemistry
Catalytic asymmetric C–H insertion reactions of vinyl carbocations
S. K. Nistanaki et al.

1092 Computer science
Competition-level code generation with AlphaCode *Y. Li* et al.
PERSPECTIVE p. 1056

REPORTS

1097 Chemical genetics
IFITM proteins assist cellular uptake of diverse linked chemotypes *K. Lou* et al.
PERSPECTIVE p. 1054

1105 Martian geology
Aqueous alteration processes in Jezero crater, Mars—implications for organic geochemistry
E. L. Scheller et al.

1111 Antibiotic resistance
Tuberculosis treatment failure associated with evolution of antibiotic resilience
Q. Liu et al.

1118 Crystallization
Liquid metal synthesis solvents for metallic crystals *S. A. Idrus-Saidi* et al.

1125 Piezoelectrics
Giant electric field–induced strain in lead-free piezoceramics
G. Huangfu et al.

DEPARTMENTS

1027 Editorial
Thank you, Tony!
By H. H. Thorp
PODCAST

1134 Working Life
Working through grief
By M. Schultz

ON THE COVER



This image shows blocks with snippets of programs of varying viability (height) created by humans (purple) or by AlphaCode (white). AlphaCode is an artificial intelligence system that writes computer programs at a competitive level and solves new problems by generating and filtering millions of diverse candidates. In the future, code generation systems like AlphaCode may assist and educate programmers. See pages 1056 and 1092. *Credit: DeepMind*

Science Staff 1026
Science Careers 1131

PHOTO: JULIA KREINER/UNIVERSITY OF BRITISH COLUMBIA

SCIENCE (ISSN 0036-8075) is published weekly on Friday, except last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005. Periodicals mail postage (publication No. 484460) paid at Washington, DC, and additional mailing offices. Copyright © 2022 by the American Association for the Advancement of Science. The title SCIENCE is a registered trademark of the AAAS. Domestic individual membership, including subscription (12 months): $165 ($74 allocated to subscription). Domestic institutional subscription (51 issues): $2212; Foreign postage extra: Air assist delivery: $98. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request, GST #125488122. Publications Mail Agreement Number 1069624. **Printed in the U.S.A.**
Change of address: Allow 4 weeks, giving old and new addresses and 8-digit account number. **Postmaster:** Send change of address to AAAS, P.O. Box 96178, Washington, DC 20090–6178. **Single-copy sales:** $15 each plus shipping and handling available from backissues.science.org; bulk rate on request. **Authorization to reproduce** material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act can be obtained through the Copyright Clearance Center (CCC), www.copyright.com. The identification code for *Science* is 0036-8075. *Science* is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

Science AAAS

Editor-in-Chief Holden Thorp, hthorp@aaas.org

Executive Editor Valda Vinson

Editor, Research Jake S. Yeston **Editor, Insights** Lisa D. Chong **Managing Editor** Lauren Kmec

DEPUTY EDITORS Gemma Alderton (UK),Stella M. Hurtley (UK), Phillip D. Szuromi, Sacha Vignieri SR. EDITORS Caroline Ash (UK), Michael A. Funk, Brent Grocholski, Pamela J. Hines, Di Jiang, Priscilla N. Kelly, Marc S. Lavine (Canada), Mattia Maroso, Yevgeniya Nusinovich, Ian S. Osborne (UK), L. Bryan Ray, Seth Thomas Scanlon (UK), H. Jesse Smith, Keith T. Smith (UK), Jelena Stajic, Peter Stern (UK), Valerie B. Thompson, Brad Wible ASSOCIATE EDITORS Bianca Lopez, Madeleine Seale (UK), Corinne Simonti, Yury V. Suleymanov, Ekeoma Uzogara LETTERS EDITOR Jennifer Sills LEAD CONTENT PRODUCTION EDITORS Chris Filiatreau, Harry Jach SR. CONTENT PRODUCTION EDITOR Amelia Beyna CONTENT PRODUCTION EDITORS Julia Haber-Katris, Nida Masiulis, Abigail Shashikanth, Suzanne M. White SR. EDITORIAL MANAGERS Carolyn Kyle, Beverly Shields SR. PROGRAM ASSOCIATE Maryrose Madrid EDITORIAL ASSOCIATE Joi S. Granger SR. EDITORIAL COORDINATORS Aneera Dobbins, Jeffrey Hearn, Lisa Johnson, Jerry Richardson, Hilary Stewart (UK), Alice Whaley (UK), Anita Wynn EDITORIAL COORDINATORS Maura Byrne, Alexander Kief, Ronmel Navas, Isabel Schnaidt, Qiyam Stewart, Brian White RESEARCH & DATA ANALYST Jessica L. Slater ADMINISTRATIVE COORDINATOR Karalee P. Rogers ASI DIRECTOR, OPERATIONS Janet Clements (UK) ASI SR. OFFICE ADMINISTRATOR Jessica Waldock (UK)

News Editor Tim Appenzeller

NEWS MANAGING EDITOR John Travis INTERNATIONAL EDITOR Martin Enserink DEPUTY NEWS EDITORS Shraddha Chakradhar, Elizabeth Culotta, Lila Guterman, David Grimm, Eric Hand (Europe), David Malakoff SR. CORRESPONDENTS Daniel Clery (UK), Jon Cohen, Jeffrey Mervis, Elizabeth Pennisi ASSOCIATE EDITORS Jeffrey Brainard, Michael Price, Kelly Servick NEWS REPORTERS Adrian Cho, Jennifer Couzin-Frankel, Jocelyn Kaiser, Rodrigo Pérez Ortega (Mexico City), Robert F. Service, Erik Stokstad, Paul Voosen, Meredith Wadman INTERNS Zack Savitsky, Viviana Flores, Katherine Irving CONTRIBUTING CORRESPONDENTS Warren Cornwall, Andrew Curry (Berlin), Ann Gibbons, Sam Kean, Eli Kintisch, Kai Kupferschmidt (Berlin), Andrew Lawler, Mitch Leslie, Eliot Marshall, Virginia Morell, Dennis Normile (Tokyo), Elisabeth Pain (Careers), Charles Piller, Gabriel Popkin, Michael Price, Joshua Sokol, Richard Stone, Emily Underwood, Gretchen Vogel (Berlin), Lizzie Wade (Mexico City) CAREERS Rachel Bernstein (Editor), Katie Langin (Associate Editor) COPY EDITORS Julia Cole (Senior Copy Editor), Morgan Everett, Cyra Master (Copy Chief) ADMINISTRATIVE SUPPORT Meagan Weiland

Creative Director Beth Rakouskas

DESIGN MANAGING EDITOR Chrystal Smith GRAPHICS MANAGING EDITOR Chris Bickel PHOTOGRAPHY MANAGING EDITOR Emily Petersen MULTIMEDIA MANAGING PRODUCER Kevin McLean WEB CONTENT STRATEGY MANAGER Kara Estelle-Powers DESIGN EDITOR Marcy Atarod DESIGNER Christina Aycock SENIOR SCIENTIFIC ILLUSTRATOR Valerie Altounian SCIENTIFIC ILLUSTRATORS Austin Fisher, Kellie Holoski, Ashley Mastin INTERACTIVE GRAPHICS EDITOR Kelly Franklin SENIOR GRAPHICS SPECIALISTS Holly Bishop, Nathalie Cary SENIOR PHOTO EDITOR Charles Borst SENIOR PODCAST PRODUCER Sarah Crespi VIDEO PRODUCER Meagan Cantwell SOCIAL MEDIA STRATEGIST Jessica Hubbard SOCIAL MEDIA PRODUCER Sabrina Jenkins WEB DESIGNER Jennie Pajerowski

Chief Executive Officer and Executive Publisher Sudip Parikh

Publisher, Science Family of Journals Bill Moran

DIRECTOR, BUSINESS SYSTEMS AND FINANCIAL ANALYSIS Randy Yi DIRECTOR, BUSINESS OPERATIONS & ANALYSIS Eric Knott DIRECTOR OF ANALYTICS Enrique Gonzales MANAGER, BUSINESS OPERATIONS Jessica Tierney MANAGER, BUSINESS ANALYSIS Cory Lipman BUSINESS ANALYSTS Kurt Ennis, Maggie Clark FINANCIAL ANALYST Isacco Fusi BUSINESS OPERATIONS ADMINISTRATOR Taylor Fisher SENIOR PRODUCTION MANAGER Jason Hillman SENIOR MANAGER, PUBLISHING AND CONTENT SYSTEMS Marcus Spiegler CONTENT OPERATIONS MANAGER Rebecca Doshi SENIOR CONTENT & PUBLISHING SYSTEMS SPECIALIST Jacob Hedrick SENIOR PRODUCTION SPECIALIST Kristin Wowk PRODUCTION SPECIALISTS Kelsey Cartelli, Audrey Diggs DIGITAL PRODUCTION MANAGER Lisa Stanford CONTENT SPECIALIST Kimberley Oster ADVERTISING PRODUCTION OPERATIONS MANAGER Deborah Tompkins DESIGNER, CUSTOM PUBLISHING Jeremy Huntsinger SR. TRAFFIC ASSOCIATE Christine Hall SPECIAL PROJECTS ASSOCIATE Sarah Dhere

ASSOCIATE DIRECTOR, BUSINESS DEVELOPMENT Justin Sawyers GLOBAL MARKETING MANAGER Allison Pritchard DIGITAL MARKETING MANAGER Aimee Aponte JOURNALS MARKETING MANAGER Shawana Arnold MARKETING ASSOCIATES Aaron Helmbrecht, Ashley Hylton, Mike Romano, Lorena Chirinos Rodriguez, Jenna Voris SENIOR DESIGNER Kim Huynh

DIRECTOR AND SENIOR EDITOR, CUSTOM PUBLISHING Sean Sanders ASSISTANT EDITOR, CUSTOM PUBLISHING Jackie Oberst PROJECT MANAGER Melissa Collins

DIRECTOR, PRODUCT & PUBLISHING DEVELOPMENT Chris Reid DIRECTOR, BUSINESS STRATEGY AND PORTFOLIO MANAGEMENT Sarah Whalen ASSOCIATE DIRECTOR, PRODUCT MANAGMENT Kris Bishop PRODUCT DEVELOPMENT MANAGER Scott Chernoff PUBLISHING TECHNOLOGY MANAGER Michael Di Natale SR. PRODUCT ASSOCIATE Robert Koepke PRODUCT ASSOCIATE Caroline Breul, Anne Mason SPJ ASSOCIATE MANAGER Samantha Bruno Fuller SPJ ASSOCIATE Casey Buchta

MARKETING MANAGER Kess Knight BUSINESS DEVELOPMENT MANAGER Rasmus Andersen SENIOR INSTITUTIONAL LICENSING MANAGER Ryan Rexroth INSTITUTIONAL LICENSING MANAGER Marco Castellan, Claudia Paulsen-Young SENIOR MANAGER, INSTITUTIONAL LICENSING OPERATIONS Judy Lillibridge SENIOR OPERATIONS ANALYST Lana Guz SYSTEMS & OPERATIONS ANALYST Ben Teincuff FULFILLMENT ANALYST Aminta Reyes

DIRECTOR, GLOBAL SALES Tracy Holmes US EAST COAST AND MID WEST SALES Stephanie O'Connor US MID WEST, MID ATLANTIC AND SOUTH EAST SALES Chris Hoag US WEST COAST SALES Lynne Stickrod ASSOCIATE DIRECTOR, ROW Roger Goncalves SALES REP, ROW Sarah Lelarge SALES ADMIN ASSISTANT, ROW Victoria Glasbey DIRECTOR OF GLOBAL COLLABORATION AND ACADEMIC PUBLISHING RELATIONS, ASIA Xiaoying Chu ASSOCIATE DIRECTOR, INTERNATIONAL COLLABORATION Grace Yao SALES MANAGER Danny Zhao MARKETING MANAGER Kilo Lan ASCA CORPORATION, JAPAN Rie Rambelli (Tokyo), Miyuki Tani (Osaka)

DIRECTOR, COPYRIGHT, LICENSING AND SPECIAL PROJECTS Emilie David RIGHTS AND PERMISSIONS ASSOCIATE Elizabeth Sandler LICENSING ASSOCIATE Virginia Warren CONTRACT SUPPORT SPECIALIST Michael Wheeler

MAIN HEADQUARTERS
Science/AAAS
1200 New York Ave. NW
Washington, DC 20005

SCIENCE INTERNATIONAL
Clarendon House
Clarendon Road
Cambridge, CB2 8FH, UK

SCIENCE CHINA
Room 1004, Culture Square
No. 59 Zhongguancun St.
Haidian District, Beijing, 100872

SCIENCE JAPAN
ASCA Corporation
Sibaura TY Bldg. 4F, 1-14-5
Shibaura Minato-ku
Tokyo, 108-0073 Japan

EDITORIAL
science_editors@aaas.org

NEWS
science_news@aaas.org

INFORMATION FOR AUTHORS
science.org/authors/science-information-authors

REPRINTS AND PERMISSIONS
science.org/help/reprints-and-permissions

MEDIA CONTACTS
scipak@aaas.org

MULTIMEDIA CONTACTS
SciencePodcast@aaas.org
ScienceVideo@aaas.org

INSTITUTIONAL SALES AND SITE LICENSES
science.org/librarian

PRODUCT ADVERTISING & CUSTOM PUBLISHING
advertising.science.org/products-services
science_advertising@aaas.org

CLASSIFIED ADVERTISING
advertising.science.org/science-careers
advertise@sciencecareers.org

JOB POSTING CUSTOMER SERVICE
employers.sciencecareers.org
support@sciencecareers.org

MEMBERSHIP AND INDIVIDUAL SUBSCRIPTIONS
science.org/subscriptions

MEMBER BENEFITS
aaas.org/membership/benefits

AAAS BOARD OF DIRECTORS
CHAIR Susan G. Amara
PRESIDENT Gilda A. Barabino
PRESIDENT-ELECT Keith Yamamoto
TREASURER Carolyn N. Ainslie
CHIEF EXECUTIVE OFFICER Sudip Parikh
BOARD Cynthia M. Beall, Ann Bostrom, Janine Austin Clayton, Kaye Husbands Fealing, Maria M. Klawe, Jane Maienschein, Robert B. Millard, Babak Parviz, William D. Provine, Juan S. Ramírez Lugo, Susan M. Rosenberg

Science serves as a forum for discussion of important issues related to the advancement of science by publishing material on which a consensus has been reached as well as including the presentation of minority or conflicting points of view. Accordingly, all articles published in *Science*—including editorials, news and comment, and book reviews—are signed and reflect the individual views of the authors and not official points of view adopted by AAAS or the institutions with which the authors are affiliated.

BOARD OF REVIEWING EDITORS (Statistics board members indicated with **S**)

Erin Adams, *U. of Chicago*
Takuzo Aida, *U. of Tokyo*
Leslie Aiello, *Wenner-Gren Fdn.*
Deji Akinwande, *UT Austin*
Judith Allen, *U. of Manchester*
Marcella Alsan, *Harvard U.*
James Analytis, *UC Berkeley*
Paola Arlotta, *Harvard U.*
Delia Baldassarri, *NYU*
Nenad Ban, *ETH Zürich*
Christopher Barratt, *U. of Dundee*
Nandita Basu, *U. of Waterloo*
Franz Bauer, *Pontificia U. Católica de Chile*
Ray H. Baughman, *UT Dallas*
Carlo Beenakker, *Leiden U.*
Yasmine Belkaid, *NIAID, NIH*
Philip Benfey, *Duke U.*
Kiros T. Berhane, *Columbia U.*
Joseph J. Berry, *NREL*
Alessandra Biffi, *Harvard Med.*
Chris Bowler, *École Normale Supérieure*
Ian Boyd, *U. of St. Andrews*
Malcolm Brenner, *Baylor Coll. of Med.*
Emily Brodsky, *UC Santa Cruz*
Ron Brookmeyer, *UCLA* (**S**)
Christian Büchel, *UKE Hamburg*
Dennis Burton, *Scripps Res.*
Carter Tribley Butts, *UC Irvine*
György Buzsáki, *NYU School of Med.*
Mariana Byndloss, *Vanderbilt U. Med. Ctr.*
Annmarie Carlton, *UC Irvine*
Simon Cauchemez, *Inst. Pasteur*
Ling-Ling Chen, *SIBCB, CAS*
Wendy Cho, *UIUC*
Ib Chorkendorff, *Denmark TU*
Karlene Cimprich, *Stanford U.*
James J. Collins, *MIT*
Robert Cook-Deegan, *Arizona State U.*
Virginia Cornish *Columbia U.*
Carolyn Coyne, *Duke U.*
Roberta Croce, *VU Amsterdam*
Christina Curtis, *Stanford U.*
Ismaila Dabo, *Penn State U.*
Jeff L. Dangl, *UNC*
Chiara Daraio, *Caltech*
Nicolas Dauphas, *U. of Chicago*
Frans de Waal, *Emory U.*
Claude Desplan, *NYU*
Sandra Díaz, *U. Nacional de Córdoba*
Samuel Díaz-Muñoz, *UC Davis*
Ulrike Diebold, *TU Wien*
Stefanie Dimmeler, *Goethe-U. Frankfurt*
Hong Ding, *Inst. of Physics, CAS*
Dennis Discher, *UPenn*
Jennifer A. Doudna, *UC Berkeley*
Ruth Drdla-Schutting, *Med. U. Vienna*
Raissa M. D'Souza, *UC Davis*
Bruce Dunn, *UCLA*
William Dunphy, *Caltech*
Scott Edwards, *Harvard U.*
Todd Ehlers, *U. of Tübingen*
Nader Engheta, *UPenn*
Karen Ersche, *U. of Cambridge*
Beate Escher, *UFZ & U. of Tübingen*
Barry Everitt, *U. of Cambridge*
Vanessa Ezenwa, *U. of Georgia*
Toren Finkel, *U. of Pitt. Med. Ctr.*
Gwenn Flowers, *Simon Fraser U.*
Natascha Förster Schreiber, *MPI Extraterrestrial Phys.*
Peter Fratzl, *MPI Potsdam*
Elaine Fuchs, *Rockefeller U.*
Jay Gallagher, *U. of Wisconsin*
Daniel Geschwind, *UCLA*
Ramon Gonzalez, *U. of South Florida*
Sandra González-Bailón, *UPenn*
Gillian Griffiths, *U. of Cambridge*
Nicolas Gruber, *ETH Zürich*
Hua Guo, *U. of New Mexico*
Taekjip Ha, *Johns Hopkins U.*
Daniel Haber, *Mass. General Hos.*
Sharon Hammes-Schiffer, *Yale U.*
Wolf-Dietrich Hardt, *ETH Zürich*
Louise Harra, *UCL*
Carl-Philipp Heisenberg, *IST Austria*
Janet G. Hering, *Eawag*
Christoph Hess, *U. of Basel & U. of Cambridge*
Heather Hickman, *NIAID, NIH*
Hans Hilgenkamp, *U. of Twente*
Janneke Hille Ris Lambers, *ETH Zürich*
Kai-Uwe Hinrichs, *U. of Bremen*
Deirdre Hollingsworth, *U. of Oxford*
Randall Hulet, *Rice U.*
Auke Ijspeert, *EPFL*
Gwyneth Ingram, *ENS Lyon*
Darrell Irvine, *MIT*
Akiko Iwasaki, *Yale U.*
Stephen Jackson, *USGS & U. of Arizona*
Erich Jarvis, *Rockefeller U.*
Peter Jonas, *IST Austria*
Johanna Joyce, *U. de Lausanne*
Matt Kaeberlein, *U. of Wash.*
William Kaelin Jr., *Dana-Farber*
Daniel Kammen, *UC Berkeley*
Kisuk Kang, *Seoul Nat. U.*
Sabine Kastner, *Princeton U.*
V. Narry Kim, *Seoul Nat. U.*
Robert Kingston, *Harvard Med.*
Nancy Knowlton, *Smithsonian Institution*
Etienne Koechlin, *École Normale Supérieure*
Alex L. Kolodkin, *Johns Hopkins U.*
LaShanda Korley, *U. of Delaware*
Julija Krupic, *U. of Cambridge*
Paul Kubes, *U. of Calgary*
Chris Kuzawa, *Northwestern U.*
Laura Lackner, *Northwestern U.*
Gabriel Lander, *Scripps Res.* (**S**)
Mitchell A. Lazar, *UPenn*
Hedwig Lee, *Duke U.*
Luis Liz-Marzán, *CIC biomaGUNE*
Omar Lizardo, *UCLA*
Jonathan Losos, *Wash. U. in St. Louis*
Ke Lu, *Inst. of Metal Res., CAS*
Christian Lüscher, *U. of Geneva*
Jean Lynch-Stieglitz, *Georgia Inst. of Tech.*
David Lyons, *U. of Edinburgh*
Fabienne Mackay, *QIMR Berghofer*
Zeynep Madak-Erdogan, *UIUC*
Anne Magurran, *U. of St. Andrews*
Ari Pekka Mähönen, *U. of Helsinki*
Asifa Majid, *U. of Oxford*
Oscar Marín, *King's Coll. London*
Charles Marshall, *UC Berkeley*
Christopher Marx, *U. of Idaho*
David Masopust, *U. of Minnesota*
Geraldine Masson, *CNRS*
C. Robertson McClung, *Dartmouth*
Rodrigo Medellín, *U. Nacional Autónoma de México*
C. Jessica Metcalf, *Princeton U.*
Baoxia Mi, *UC Berkeley*
Tom Misteli, *NCI, NIH*
Alison Motsinger-Reif, *NIEHS, NIH* (**S**)
Danielle Navarro, *U. of New South Wales*
Daniel Nettle, *Newcastle U.*
Daniel Neumark, *UC Berkeley*
Thi Hoang Duong Nguyen, *MRC LMB*
Beatriz Noheda, *U. of Groningen*
Helga Nowotny, *Vienna Sci. & Tech. Fund*
Pilar Ossorio, *U. of Wisconsin*
Andrew Oswald, *U. of Warwick*
Isabella Pagano, *Istituto Nazionale di Astrofisica*
Elizabeth Levy Paluck, *Princeton U.*
Jane Parker, *MPI Cologne*
Giovanni Parmigiani, *Dana-Farber* (**S**)
Daniel Pauly, *U. of British Columbia*
Ana Pêgo, *U. do Porto*
Samuel Pfaff, *Salk Inst.*
Julie Pfeiffer, *UT Southwestern Med. Ctr.*
Philip Phillips, *UIUC*
Matthieu Piel, *Inst. Curie*
Kathrin Plath, *UCLA*
Martin Plenio, *Ulm U.*
Katherine Pollard, *UCSF*
Elvira Poloczanska, *Alfred-Wegener-Inst.*
Julia Pongratz, *Ludwig Maximilians U.*
Philippe Poulin, *CNRS*
Lei Stanley Qi, *Stanford U.*
Trevor Robbins, *U. of Cambridge*
Joeri Rogelj, *Imperial Coll. London*
Amy Rosenzweig, *Northwestern U.*
Mike Ryan, *UT Austin*
Miquel Salmeron, *Lawrence Berkeley Nat. Lab*
Nitin Samarth, *Penn State U.*
Erica Ollmann Saphire, *La Jolla Inst.*
Joachim Saur, *U. zu Köln*
Alexander Schier, *Harvard U.*
Wolfram Schlenker, *Columbia U.*
Susannah Scott, *UC Santa Barbara*
Anuj Shah, *U. of Chicago*
Vladimir Shalaev, *Purdue U.*
Jie Shan, *Cornell U.*
Beth Shapiro, *UC Santa Cruz*
Jay Shendure, *U. of Wash.*
Steve Sherwood, *U. of New South Wales*
Brian Shoichet, *UCSF*
Robert Siliciano, *JHU School of Med.*
Lucia Sivilotti, *UCL*
Emma Slack, *ETH Zürich & U. of Oxford*
Richard Smith, *UNC* (**S**)
John Speakman, *U. of Aberdeen*
Tara Spires-Jones, *U. of Edinburgh*
Allan C. Spradling, *Carnegie Institution for Sci.*
V. S. Subrahmanian, *Northwestern U.*
Ira Tabas, *Columbia U.*
Eriko Takano, *U. of Manchester*
A. Alec Talin, *Sandia Natl. Labs*
Patrick Tan, *Duke-NUS Med. School*
Sarah Teichmann, *Wellcome Sanger Inst.*
Rocio Titiunik, *Princeton U.*
Shubha Tole, *Tata Inst. of Fundamental Res.*
Maria-Elena Torres Padilla, *Helmholtz Zentrum München*
Kimani Toussaint, *Brown U.*
Barbara Treutlein, *ETH Zürich*
Jason Tylianakis, *U. of Canterbury*
Wim van der Putten, *Netherlands Inst. of Ecology*
Matthew Vander Heiden, *MIT*
Ivo Vankelecom, *KU Leuven*
Judith Varner, *UC San Diego*
Henrique Veiga-Fernandes, *Champalimaud Fdn.*
Reinhilde Veugelers, *KU Leuven*
Bert Vogelstein, *Johns Hopkins U.*
Julia Von Blume, *Yale School of Med.*
David Wallach, *Weizmann Inst.*
Jane-Ling Wang, *UC Davis* (**S**)
Jessica Ware, *Amer. Mus. of Natural Hist.*
David Waxman, *Fudan U.*
Chris Wikle, *U. of Missouri* (**S**)
Terrie Williams, *UC Santa Cruz*
Ian A. Wilson, *Scripps Res.* (**S**)
Hao Wu, *Harvard U.*
Li Wu, *Tsinghua U.*
Wei Xie, *Tsinghua U.*
Benjamin Youngblood, *St. Jude*
Yu Xie, *Princeton U.*
Jan Zaanen, *Leiden U.*
Kenneth Zaret, *UPenn School of Med.*
Lidong Zhao, *Beihang U.*
Bing Zhu, *Inst. of Biophysics, CAS*
Xiaowei Zhuang, *Harvard U.*
Maria Zuber, *MIT*

EDITORIAL

Thank you, Tony!

Anthony Fauci—"Tony" to friends, colleagues, and many journalists—has never backed down from controversy. During the worst of the AIDS crisis, when HIV infection was a death sentence, he was targeted by protestors frustrated by the US government's sluggish response. Instead of ignoring their complaints, Fauci listened, acted, and eventually came to be seen as an ally to activists, all the while battling charlatans who tried to insist that HIV didn't cause AIDS. When the COVID-19 pandemic came along, the long-time director of the National Institute of Allergy and Infectious Diseases was a natural choice to help lead the country's response, and he soon became one of the most recognizable physician-scientists in the world. Fauci recently announced that he would end his time in the federal government in December 2022. He sat down with me to talk about the challenges facing science and his plans for the future. The full transcript of our interview, which is filled with valuable insights into science policy and communication, is posted on the Editor's Blog, but I wanted to share a few highlights here.

We talked a lot about misinformation and the forces sowing doubts about COVID-19 vaccines and interventions such as masks and lockdowns. When I asked him how this round of science denial compared with what he saw during the HIV/AIDS crisis, he agreed that the two shared common themes, but argued that our current crisis is fundamentally different, due largely to powerful platforms offered to antiscience actors by social media and other channels. "I always say, 'The best way to counter misinformation is to flood the system with correct information,'" he told me. "That's still true, but you're fighting against a big surge of misinformation."

We agreed that one of the biggest challenges with which the scientific community must grapple is misinformation that comes from within. "I ran into that like a stone wall at a congressional hearing," he said, describing how his skepticism of an argument put forward by researchers who questioned the necessity of COVID-19 lockdowns was met with admonishment. "You're arrogant because you're disrespecting other scientists," he was told. Fauci's response: "You don't want to be being disparaging of anyone, but you've got to come back and push back against things that are not true."

We also talked about how we need to do a better job of conveying that science is a process and not a collection of facts set in stone. I was trying to stay away from the pushback Fauci received each time pandemic guidance had to be revised, but it hung in the air, unspoken. "For those who don't understand the iterative nature of science, you are flip-flopping, and you've undermined the entire scientific process," he observed. "What we can do better is articulate to people that when we tell them something based on the evidence we have now, that we are dealing with an evolving situation."

> "As a scientist, you have an obligation…to speak the truth and not be evasive…"
> **Anthony Fauci,**
> National Institute of Allergy and Infectious Diseases

Unlike many of his government peers, Fauci has always been forthcoming with the media. "As a scientist, you have an obligation…to speak the truth and not be evasive, because that's antithetical to the scientific principles," he told me. "And when I talk to my scientific colleagues in government, often they're afraid to say something, and I'm telling them behind the scenes, 'But what's the problem? You're not going to lose your job for telling the truth.' And I always found that that has served me very, very well." In my opinion, his preference for straight talk over carefully crafted statements has also served the public well.

When I asked what was next, Fauci was scrupulous about the fact that it couldn't be worked out until he left government, but he gave me the broad strokes. "I want to be able to write. I want to be able to lecture. I want to be able to advise when asked for advice," he said. "And [I want to] use the benefit of my experience to do two things: to help people and organizations, including the federal government, or universities, who could benefit from my experience, but also, as important, to inspire young people to either go into science, or if they are in science, to pursue it in a manner that benefits the public health and to do it with public service."

–H. Holden Thorp



H. Holden Thorp
Editor-in-Chief, *Science* journals.
hthorp@aaas.org; @hholdenthorp

PHOTO: CAMERON DAVIDSON

10.1126/science.adg0283



NEWS

IN BRIEF

Edited by **David Malakoff**

CLIMATE SCIENCE

Volcano and NASA deliver blows to climate monitoring

An erupting fissure high on Mauna Loa's Northeast Rift Zone.

Efforts to monitor global greenhouse gas emissions suffered two setbacks last week—one by chance, one by choice. In Hawaii, the first eruption of the Mauna Loa volcano since 1984 has cut off road access and power to a famed summit lab that has monitored atmospheric carbon dioxide (CO_2) levels since 1958. Although lava flows have so far spared the lab, which is run by the National Oceanic and Atmospheric Administration (NOAA), measurements are unlikely to resume for several months. That means tracking data will have to be sustained by other NOAA-run stations around the world. NASA, meanwhile, canceled development of its troubled GeoCarb mission, designed to track CO_2 levels from geostationary orbit. The mission's price tag had ballooned from $171 million to more than $600 million since 2016. NASA said new technology, including a hyperspectral instrument mounted on the International Space Station this year, has created alternatives to GeoCarb and it will begin work later this decade on a new greenhouse gas monitoring satellite.

Antiparasite pill impresses

NEGLECTED DISEASES | A new drug that can treat human African trypanosomiasis—commonly known as sleeping sickness—with just one dose has shown promise in a clinical trial in the Democratic Republic of the Congo and Guinea. The rare disease is caused by the parasite *Trypanosoma brucei gambiense*, which is transmitted by the tsetse fly. Left untreated, it is deadly. Until a few years ago, treatment required hospitalization and a series of injected drugs—difficult to administer in the remote regions where the disease occurs. In 2019, African nations began using a treatment that required 10 daily pills. However, the new drug, called acoziborole, could be even simpler. In a trial in 208 patients, the one-dose pill cured 95% of the people treated, even if they were suffering from late-stage disease. Researchers are now conducting a placebo-controlled trial with 900 participants to collect more safety data.

Avian flu strikes South America

EPIDEMIOLOGY | Migratory birds in South America are now spreading the highly pathogenic H5N1 avian flu virus that has already led to the deaths of record numbers of poultry and wild birds in Europe and North America. Last week, Peru, Ecuador, and Venezuela declared agricultural emergencies and instructed farmers to slaughter poultry at infected farms. In Colombia, H5N1 was found last month in a backyard flock. Galápagos National Park is monitoring wild birds for infections. In Peru, more than 10,000 pelicans have died of the extremely contagious virus. Meanwhile, the agriculture ministry in Mexico—where commercial farms have slaughtered nearly 4 million birds since October—announced more than 41 million doses of poultry vaccine have been delivered to farmers in high-risk regions.

A shortcut for particle physics

MATH | For decades, theoretical particle physicists have struggled with vexing calculus problems called Feynman integrals, which are central to nearly every calculation of how subatomic particles interact. Now, theorists in China have found a general numerical method to solve the integrals much more quickly. The approach transforms them into differential equations, which can be solved easily by a computer—although the method depends on a key new insight and lots and

PHOTO: L. GALLANT/U.S. GEOLOGICAL SURVEY

lots of linear algebra. Published last week in *Physical Review Letters*, the method should be widely useful, including in making predictions to be tested at the world's biggest atom smasher, Europe's Large Hadron Collider. A software package deploying the method is already getting heavy use, developers say.

Huge telescope to get underway

RADIO ASTRONOMY | Construction will soon begin on the world's biggest radio telescope, the Square Kilometre Array Observatory. Officials at the twin sites in South Africa and Western Australia announced contracts to install foundations, power, and fiber connections for the 131,072 Christmas tree–shaped antennas that will detect low frequencies and the 133 dishes to monitor higher frequencies. The dishes will be added to the existing 64 dishes of South Africa's MeerKAT array in a configuration extending across hundreds of square kilometers. The €2 billion facility, to be completed by the end of the decade, will probe the universe's first stars and galaxies, study cosmic magnetism and gravity, and listen for alien civilizations.

Daunting biodiversity talks

CONSERVATION | Delegates who arrived in Montreal this week for a key summit meeting on conserving global biodiversity face a heavy workload. Parties to the Convention for Biological Diversity must finalize a 261-page document that contains some 1800 words and phrases in brackets because negotiators have yet to reach agreement on pollution control targets, agricultural reforms, and other issues. The meeting is scheduled to end on 19 December.

Meet a digital Blue Marble

CLIMATE SCIENCE | Fifty years ago this week, astronauts aboard Apollo 17, NASA's last crewed mission to the Moon, took an iconic photograph that became known as the Blue Marble—the first color photo taken by a person that shows Earth's full day side. This week, researchers led by the Max Planck Institute for Meteorology re-created that image during a test run of a cutting-edge digital climate model. The model can simulate climatic phenomena, such as storms and ocean eddies, at 1-kilometer resolution, as much as 100 times sharper than typical global simulations. To duplicate the swirling winds of the Blue Marble—including a cyclone over the Indian Ocean—the researchers fed spotty weather records from 1972 into the



A climate model created this replica of the iconic 1972 Blue Marble photo of Earth.

supercomputer-powered software, which can handle time spans of just a few days. Researchers say the stunt highlights the growing sophistication of high-resolution climate models, which are expected to form the core of the European Union's Destination Earth project, which aims to create a "digital twin" of Earth (*Science*, 2 October 2020, p. 16).

NASA honors UV researcher

ATMOSPHERIC SCIENCE | NASA said last week it is renaming an atmospheric science probe known as GLIDE after the pioneering astronomer George Carruthers. The Carruthers Geocorona Observatory, set for launch in 2025, will study Earth's upper atmosphere, where neutral hydrogen gas scatters far-ultraviolet (far-UV) radiation from the Sun. Carruthers, who died in 2020, pioneered UV astronomy in the 1960s, building cameras that were lofted above the UV-blocking atmosphere with sounding rockets to get brief glimpses of the sky. The instruments showed that molecular hydrogen exists in the voids between stars. Carruthers also built a camera that, in 1972, was operated on the Moon by Apollo 16 astronauts, taking the first UV images of Earth's geocorona. He later snapped Halley's Comet in UV and built a camera for the Space Shuttle. The decision appears to mark the first time NASA has named a space mission after a Black scientist. Carruthers also worked to promote and mentor young Black researchers and students interested in careers in science.



A stone panel found in Turkey depicts people interacting with fearsome wild animals.

ARCHAEOLOGY

A Neolithic narrative?

A stone panel carved some 10,500 years ago by Neolithic hunter-gatherers in what is now southern Turkey is the region's oldest known example of a story told in art, an archaeologist claimed this week. Discovered in 2021 in the village of Sayburç, the 3.7-meter-long panel portrays two human figures encountering wild animals. One, with six fingers, crouches beside a sharp-horned bull. The other, flanked by two snarling leopards, clutches an erect penis. Researchers caution it is very difficult to know whether the ancient artists intended to tell a story—and what it might be. But unlike other Neolithic carvings from the region, this panel shows characters interacting and appears to have "narrative integrity," archaeologist Eylem Özdoğan of Istanbul University wrote in *Antiquity*, including "a progressing scene [that] ... suggests that one or more related events or stories are being told."

PHOTOS: (TOP TO BOTTOM) MPI, DKRZ, AND NVIDIA/NASA; SAYBURÇ PROJECT ARCHIVE



Despite safety issues, an antibody that appears to slow Alzheimer's disease could offer hope to patients with few options.

PHOTO: CHEN ZHONGQIU/VCG/GETTY IMAGES

BIOMEDICINE

Alzheimer's drug stirs excitement—and concerns

Antibody slows cognitive decline, but deaths, brain bleeds, and swelling mar results

By **Jennifer Couzin-Frankel** *and* **Charles Piller**

A festive air pervaded an Alzheimer's disease conference last week in San Francisco, where attendees cheered what several described as a monumental breakthrough: the first treatment to clearly, albeit modestly, slow the disease's seemingly inexorable cognitive decline. In so doing, some scientists say, it substantiated the long-held hypothesis that the brain protein known as beta amyloid is a major driver of the illness. The news is "a celebration for all of us in the field," said neuroscientist Maria Carrillo, chief science officer of the Alzheimer's Association, at the meeting.

But the buoyant mood was tempered by doubts about how well the experimental therapy, an antibody called lecanemab, worked in its pivotal clinical trial—especially in people who carry two copies of *APOE4*, a gene variant that predisposes them to Alzheimer's. Safety was also a concern, with scientists debating the treatment's risks of brain swelling and bleeding. Some have linked the drug to the deaths of two trial participants and to serious brain injuries in others, although the Japanese biotech Eisai Co., lead sponsor of the clinical trial, suggested lecanemab was not to blame.

Dennis Selkoe, a Harvard University neuroscientist and a longtime supporter of the amyloid hypothesis, says he has waited decades for this moment. "For the first time in my career, I see objective evidence" that reducing amyloid in the brain produces better cognitive outcomes. "I'm on the right side of history." Besides vindicating the amyloid hypothesis, Selkoe says, the trial results herald a promising treatment. We don't "need perfection to offer this to patients."

Others strongly disagreed with both conclusions. "I do not think the benefits seen in this trial clearly outweigh the risks and despite the general excitement around a possible new treatment, I will be advising my patients to keep waiting," Matthew Schrag, a Vanderbilt University physician and neuroscientist, tweeted last week.

Eisai and its partner Biogen designed lecanemab to remove certain forms of beta amyloid, which clumps between brain cells in people with Alzheimer's and is widely thought to cause the disease's memory-robbing neurodegeneration. Other experimental therapies, including antibodies, have removed amyloids, but patients showed little sign of benefiting from those earlier drugs.

In presentations at last week's Clinical Trials on Alzheimer's Disease (CTAD) conference, along with a paper in *The New England Journal of Medicine*, the companies and several researchers expanded on a September press release that briefly detailed the results of the phase 3 lecanemab trial, which enrolled 1795 early-stage Alzheimer's patients. The talks and paper confirmed that lecanemab, given by intravenous infusion every other week, on average slowed the patients' rate of cognitive decline by 27% after 18 months, compared with people on a placebo. That corresponded to a 0.45-point separation on the 18-point dementia scale used to assess participants.

It's a "highly statistically significant" difference, said Christopher van Dyck, director of the Alzheimer's Disease Research Unit at Yale University and a study leader. The antibody also thoroughly mopped up beta amyloid, according to brain scans.

But neurologists disagree on whether the reported slowing in cognitive decline would be perceptible to many patients or caregivers. And they are keen to see whether the drug maintains its benefits. "What happens in the next 18 months?" David Knopman, a neurologist at the Mayo Clinic, said at a CTAD panel discussion. "Will the effect expand, which is what we hope, will it stabilize, or will it decline?"

Knopman and other scientists also point out that not all participants appeared to benefit equally. Women and people under age 65 showed no significant boost from the antibody, for example. The subgroups analyzed were not large enough to draw firm statistical conclusions, Selkoe counters.

The randomized portion of the lec-

anemab trial ended last year, and both the placebo and treated groups were then offered the chance to join an "extension" trial in which all received lecanemab. Together, the trials are helping bring the therapy's risks into sharper focus. Like other antibodies that target amyloids, lecanemab sometimes produced a form of brain swelling or bleeding, called amyloid-related imaging abnormalities (ARIA). Routine MRI scans showed roughly 21% of those on lecanemab had such side effects, compared with just over 9% of those on placebo. Many did not notice any symptoms.

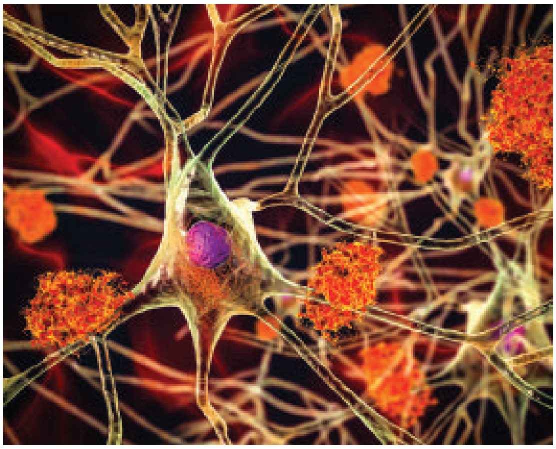
Still, 3.2% of 898 participants on the antibody in the recent trial developed ARIA with symptoms, including headache, confusion, and visual problems. (In the placebo group, 0.2% experienced this.) Schrag and others fear the bleeding risk is especially acute for Alzheimer's patients also taking anticoagulants, which interfere with clotting and are often prescribed to older people for various health problems. (Eisai declined to disclose full data on ARIA in participants who took anticoagulants while on lecanemab.)

Adding to the concern, two patients in the extension trial died following brain hemorrhages; both were on anticoagulants. One was a 65-year-old physically active woman with minimal cognitive decline. In an exclusive interview with *Science* recently, the woman's spouse described the catastrophe that unfolded after she arrived at Northwestern University Medical Center with a suspected stroke. Doctors gave her a common stroke treatment, the powerful clot-busting medication tissue plasminogen activator (tPA). In short order, she suffered massive bleeding throughout her brain's outer layer, according to an unpublished case report by Northwestern physicians, obtained by *Science*. She died a few days later without regaining consciousness.

"There's zero doubt in my mind that this is a treatment-caused illness and death. If the patient hadn't been on lecanemab she would be alive today," says Rudolph Castellani, a Northwestern neuropathologist who studies Alzheimer's and conducted the patient's autopsy at her husband's request. Although tPA does carry a risk of brain bleeds, including fatal ones on rare occasions, independent experts who reviewed the unpublished paper for *Science* said lecanemab likely played a key role in the hemorrhage and death. (Castellani, whom the husband authorized to speak to *Science* about his wife's case, says his comments reflect his personal views, not those of Northwestern.)

At last week's meeting, Marwan Sabbagh of the Barrow Neurological Institute presented further data on the trial's adverse events, including brain "macrohemorrhages" more than 1 centimeter in diameter. Such bleeding occurred in five out of 140 people—or 3.6%—on both lecanemab and anticoagulants, a far higher rate than in people only getting the antibody. One of those five was the woman who had the stroke; a second was an 87-year-old man whose death was reported in October by *STAT*. A third macrohemorrhage patient using anticoagulants experienced a devastating brain injury, according to a September paper by French researchers involved in the trial.

Besides taking anticoagulants, both people who died turned out to have cerebral amyloid angiopathy (CAA), a condition in which amyloid builds up around brain blood vessels. In attacking amyloid plaques, antiamyloid antibodies also strip away the amyloid around blood vessels—and, in doing so, can weaken them, increasing the risk of bleeds. CAA can be hard to diagnose before death and afflicts about half of Alzheimer's patients—particularly those who carry two copies of *APOE4*.



A new antibody targets the protein beta amyloid, which forms clumps (orange in this artist's concept) between brain cells.

At the CTAD conference, Sabbagh disputed that lecanemab has precipitated fatal brain injuries, suggesting the deaths were caused by tPA complications in the woman and a heart condition in the older man. But Nicolas Villain, a neurologist at Sorbonne University and an investigator on the lecanemab trial, is more uneasy. He was on the French team that described severe brain injuries—a macrohemorrhage and a serious brain swelling—in two participants on anticoagulants and urges great caution when mixing lecanemab and blood thinners.

Villain especially worries that Alzheimer's patients who have strokes while on lecanemab could die if treated with tPA, like the 65-year-old woman. Regulators "should take this case report seriously into account," agreed Andreas Charidimou, a neuroscientist at Boston University. "When there's so many unknowns it's better to be more conservative."

Pinning down the causes of deaths and severe side effects in trials can be difficult, some scientists stressed. Others argued that people with early Alzheimer's should be allowed to make their own choices about risks and benefits. "If you ask patients what risk they're willing to take with this disease, you may be surprised," Sharon Cohen, a neurologist at the Toronto Memory Program and investigator on the lecanemab trial, said during a press conference last week.

Amid this debate, lecanemab appears on the cusp of being greenlit by the U.S. Food and Drug Administration (FDA). The agency approved aducanumab, another antiamyloid antibody, last year on weaker evidence of cognitive benefits despite the complexities and costs of the treatment. Such therapies, lecanemab included, must be given intravenously, could require periodic brain scans for safety, and are likely to cost tens of thousands of dollars a year. The agency is expected to rule on lecanemab by 6 January 2023.

If it approves the antibody, FDA should at the very least require a warning against it being "given concurrently with anticoagulation or other significant blood thinners," Schrag says. He and some other scientists worry especially about people with two copies of *APOE4*. In the trial, those participants both benefited less from the treatment and faced higher risks: 9.2% of people with two copies of the gene variant showed symptomatic brain swelling, compared with 1.4% of people with no copies. Other studies of antiamyloid antibodies have also seen signs of higher risk in *APOE4* carriers.

That's a red flag even to some fans of Eisai's drug. Lecanemab is "truly a breakthrough" and "I do look forward to prescribing it," says Jason Karlawish, codirector of the Penn Memory Center at the University of Pennsylvania. But, he adds, "I will explain the benefits of *APOE* testing as a way to help a patient and family decide whether this is the right drug for them."

Karlawish says FDA should seek guidance on lecanemab from an advisory committee. Currently, no such meeting is scheduled. ■

Charles Piller's reporting was supported by the *Science* Fund for Investigative Reporting.

IMAGE: KATERYNA KON/SCIENCE SOURCE



A 2019 image of the Baltic Sea captured blooms of green algae caught up in whirling eddies.

OCEANOGRAPHY

NASA radar altimetry mission to study hidden ocean swirls

Enhanced resolution of SWOT satellite will highlight how small eddies soak up heat and carbon

By **Paul Voosen**

Eddies have been overlooked for too long. These turbulent swirls of water, ranging in size from a few kilometers to hundreds of kilometers across, peel off large ocean currents and mix heat and carbon dioxide into deeper ocean layers, like cream stirred into coffee. They are the most energetic feature of the ocean, critical to getting climate models right—but also largely invisible to satellites, except when they happen to sweep up a massive bloom of green phytoplankton.

No longer. Eddies and, on land, the ebb and flow of rivers and lakes will snap into focus after the launch of the Surface Water and Ocean Topography (SWOT) satellite, a joint venture between NASA and CNES, the French space agency. Expected to launch no earlier than 15 December from Vandenberg Space Force Base in California, on a SpaceX Falcon 9 rocket, the $1.2 billion satellite carries an altimeter that can measure the height of a water surface to within a couple centimeters, allowing researchers to infer the movements that sculpt it. "The change we expect from SWOT will be quite dramatic," says J. Thomas Farrar, a physical oceanographer at the Woods Hole Oceanographic Institution (WHOI).

For oceanographers it will be like slipping on a pair of eyeglasses, says Rosemary Morrow, a physical oceanographer at the Laboratory of Space, Geophysical, and Oceanographic Studies in Toulouse, France. The satellite will capture eddies as small as 7 kilometers across and cover nearly the entire globe every 21 days. On land, SWOT will be able to map the changing height of more than 6 million lakes, from the Great Lakes down to ponds, while also capturing flows in rivers wider than 100 meters. It will replace spotty, infrequent measurements from the ground and make the field of hydrology far more empirical, and global, than it ever has been. "It's going to help us constrain how the water cycle works in the Arctic, Africa—places where we don't have on the ground data," says Tamlin Pavelsky, a hydrologist at the University of North Carolina, Chapel Hill, and co-lead of SWOT's freshwater science team.

For nearly 4 decades, NASA and CNES have launched a series of radar altimeter satellites, which use reflected pulses of radar to measure water height. Those instruments have monitored the accelerating rise of global sea levels, a basic indicator of climate change. By measuring the ocean's bulges and dimples, they also track the large-scale currents that sweep water around the planet. But the satellites' coarse spatial resolution meant rivers and small eddies were out of reach.

SWOT gains a sharper view with the help of two 5-meter booms, each bearing an antenna to catch reflections of the radar signal SWOT pulses to Earth's surface. The widely separated antennas give SWOT the resolution to measure the height of a patch of water just kilometers wide, rather than hundreds of kilometers, bringing small eddies into view.

Armed with the precise observations, hydrologists will be able to say how lakes and rivers change seasonally, and how short-term climatic drivers, such as El Niño, affect those rhythms. For marine ecologists, SWOT will be able to chart how the levels of the world's major rivers drop each time a dam or weir interrupts them, and how severely that fragments aquatic habitats. It will also see the ripples that betray a river's shallows and deep pools, a boon to studies of how rivers evolve. And SWOT will capture flood waters as they move down river, which should help flood modelers, although the measurements won't be fast enough to help communities prepare.

Although SWOT is supposed to operate for just 3 years, its science team plans to look for correlations between the water flows it detects and features the ongoing Landsat missions see in visible light, such as changing lake and river widths. Those visible changes could serve as proxies for water levels, allowing researchers to continue to keep tabs on the planet's flows, Pavelsky says. "Even once SWOT is gone, you can still continue that analysis."

SWOT's view of eddies may be its biggest payoff. For example, it will test predictions that thousands of small eddies stir the ocean at any one time, says Sylvia Cole, a physical oceanographer at WHOI. Eddies just a few kilometers across likely play a critical role in stirring heat and carbon into the oceans near the poles, Morrow says. They also drive the mixing in smaller seas, she says. "We're probably underestimating the energy in the Mediterranean by 90% because we're missing these smaller scale structures."

At the coastlines, SWOT will provide a detailed picture of how hot spots of sea level rise in the open ocean influence coastal inundation, says Sönke Dangendorf, a physical oceanographer at Tulane University. It will also study another potential threat to coasts: small eddies that might warm nearby waters, fueling stronger hurricanes, Morrow says. "We're trapping more heat not just at the surface, but also at depth." These are important questions to answer, fast, as much of humanity lives along coasts, she adds. "Everything is felt more keenly in the coastal zone." ■

PHOTO: EUROPEAN SPACE AGENCY

SCIENTIFIC INTEGRITY

Image problems besiege Stanford president

Marc Tessier-Lavigne's early papers are subject of school and journal investigations

By **Jocelyn Kaiser**

Allegations of scientific image manipulation are threatening Stanford University President Marc Tessier-Lavigne, a neurobiologist and former biotech leader. Tessier-Lavigne came under fire last week after an investigation by the school's student newspaper revived long-standing concerns about several publications on which he was a co-author. In response, Stanford announced a misconduct investigation centering on "certain scientific articles," including two papers in *Science* and one in *Cell* dating back more than 20 years.

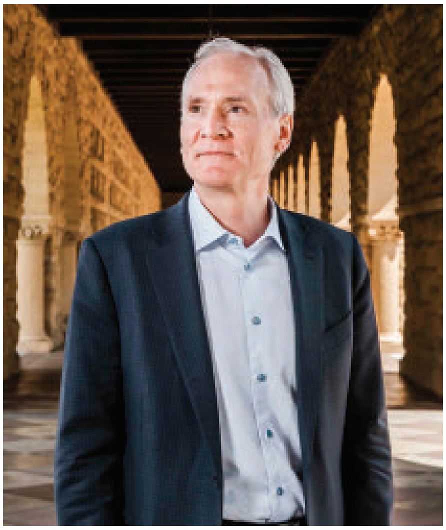
The inquiry has rattled the field of neuroscience, where Tessier-Lavigne is a respected figure, as well as the broader academic community, where some question Stanford's handling of the matter. The university's board of trustees, on which Tessier-Lavigne sits, designated five other members to lead the inquiry. One stepped aside after the paper, *The Stanford Daily*, revealed his investment firm has a $18 million stake in Denali Therapeutics, a company Tessier-Lavigne cofounded.

The initial controversy triggered scrutiny of other papers Tessier-Lavigne co-authored and revealed additional problematic images. Image manipulation experts say at least some of the errors appear minor. Still, Tessier-Lavigne should step aside during the probe, one observer suggests. "The in-house investigation will have more credibility," says Richard Smith, a former editor of *The BMJ*. "And his position as president is compromised when facing these serious accusations."

In a 5 December letter to faculty, Tessier-Lavigne said, "I welcome" the review and that "the integrity of my work is of paramount importance to me, and I take any concerns that are expressed very seriously." Although he took responsibility for any paper on which he was involved, Tessier-Lavigne noted that he was senior author on just the three papers in *Science* and *Cell* and that flagged images in papers in other journals, such as *Nature* and *The EMBO Journal*, for example, were prepared by collaborators' labs. Indeed, co-authors elsewhere have already taken responsibility for errors in various papers.

Some outsiders think Tessier-Lavigne's scientific reputation will remain intact. "A lot of important and very solid work came out of Marc's lab, as well as inevitably some studies that seemed a little overhyped. I was relieved when I clicked on the links to see that it was not some of Marc's more seminal papers that were attracting such scrutiny," says neurobiologist Barry Dickson of the Queensland Brain Institute. The *Science* and *Cell* papers "comprise a rather small fraction of the overall body of work … for which Tessier-Lavigne is well known and widely admired," adds Harvard University neurobiologist Joshua Sanes.

Tessier-Lavigne, 62, a former president of Rockefeller University who was also once chief scientific officer at Genentech, is known for groundbreaking research in the 1990s on proteins dubbed netrins. He and others showed the molecules govern the growth of nerve fibers called axons in the developing spinal cord, work for which he and two colleagues won the prestigious 2020 Gruber Prize in Neuroscience. The 1999 *Cell* paper and two 2001 *Science* publications now under scrutiny, produced when he was at the University of California, San Francisco (UCSF), came after his initial landmark netrin discoveries.



Marc Tessier-Lavigne's research on guidance cues for growing nerve fibers has drawn new scrutiny.

PHOTO: CAROLYN FONG/REDUX

Questions about certain images in the papers first surfaced 7 years ago on PubPeer, an online forum where scientists identify and discuss possible problems in published research, often anonymously. Some posts suggested various Western blots, which document a protein's presence in a sample, were repeated in more than one figure or altered. In other cases, images or parts of images appear to have been cut and pasted into other figures.

Elisabeth Bik, a specialist in spotting manipulated images, reexamined the three papers and other Tessier-Lavigne publications noted in PubPeer at the request of the *Daily*. She calls some image changes "beautification" that did not affect the papers' conclusions but describes others as "more serious."

Stanford told the *Daily* it discussed concerns about the papers with Tessier-Lavigne when he was under consideration for university president in 2015. He submitted corrections for both *Science* papers, but the journal never posted them "due to an error," said Holden Thorp, editor-in-chief of the *Science* family of journals. Tessier-Lavigne says another error has since been identified in one of the *Science* papers; Thorp says the Stanford president has been "very collaborative" in discussing possible solutions.

Tessier-Lavigne also reported errors in the *Cell* paper in 2015, but at the time its editors found no correction was necessary, the publisher Cell Press said in a statement. Last week, Bik found another manipulated image in that paper, which she says appears to be "deliberately changing the results" and could indicate data falsification. Cell Press said the "new concerns … warrant a closer look" and it now plans to investigate. The only authors on all three challenged papers are Tessier-Lavigne and a postdoc then at UCSF who apparently hasn't published a paper since 2012. She did not respond to emails from *Science*.

Some scientists pointed out that regardless of the source of the alterations in the three papers, Tessier-Lavigne was the leader of the work. He was "responsible for the integrity of the data," Bik says. All problematic papers that Tessier-Lavigne co-authored should be reviewed, she adds. "This investigation should be much wider [involving] all his collaborators' institutions."

Tessier-Lavigne's troubles are a warning to investigators everywhere to keep an eye on how lab members edit images of data, some say. "In this day and age, it's very easy to manipulate digital images. There's a lot of trust built into science and trust can be abused," says neurobiologist Tim Kennedy of McGill University. "On the flip side, science is a self-correcting enterprise." ■

COVID-19

In China, 'zero COVID' has become a Catch-22

Population chafes at control measures, but lifting them now would carry huge risks

By **Dennis Normile**

Surprised and stung by protests against draconian "zero-COVID" policies, Chinese authorities are gingerly moving to ease the burden of lockdowns, quarantines, and constant testing. But 3 years into the pandemic, China shows no sign of planning a major course change. Mathematical models suggest why: The country is still ill-prepared for living with SARS-CoV-2. Easing restrictions today would likely trigger a massive wave of infections, overwhelm health care facilities, and bring a high death toll.

"China has not achieved high vaccination rates, has not used the best type of vaccines, and it has been very slow in communicating [to the public] the eventual need to transition from elimination to suppression and mitigation," says public health scientist Nick Wilson of the University of Otago, Wellington, in New Zealand. Other countries that initially followed the zero-COVID strategy, including New Zealand, used it to buy time to ramp up vaccination rates, stockpile antivirals, and boost intensive care capacity.

The fierce protests have triggered some changes. Several provinces have started to allow people to enter public transportation, restaurants, and shopping centers without proof of a negative COVID-19 test, for example, and some close contacts of patients will be allowed to isolate at home instead of being sent to quarantine centers. Those measures are "a small but important step, and I believe more steps will follow," says a Chinese scientist who asked to remain anonymous. "This will alleviate some of the pandemic fatigue and reduce the burden on the economy and society."

But models show why the Chinese government still wants to keep a lid on. A study based on vaccination rates in March, published in *Nature Medicine* in May, found lifting zero-COVID restrictions at that point could "generate a tsunami of COVID-19 cases" over a 6-month period, with 112 million symptomatic cases, 2.7 million intensive care unit (ICU) admissions, and 1.6 million deaths. Peak demand for ICU beds would hit 1 million, more than 15 times the current capacity.

The unvaccinated would account for 77% of the fatalities, according to the authors, primarily at Fudan University. Boosting vaccination rates could slash the toll, but China's elderly population has remained wary of vaccination. Even today, only 66% of those ages 80 and older have received two doses—versus 90% of the population as a whole—and just 40% have taken boosters.

Hong Kong provides a cautionary tale: A big Omicron outbreak early this year caused nearly 6000 deaths, 96% of them in people 60 or older. At the time, Hong Kong's vaccination rates were even lower than the mainland's. In the first 3 months of this year, the city had a COVID-19 death rate of 37.7 per million population, among the highest anywhere.

A new study released on 28 November by Airfinity, a London-based health analytics firm, suggests mainland China remains vulnerable. Lifting zero COVID now could cause between 167 million and 279 million cases and between 1.3 million and 2.1 million deaths over 83 days, according to the report.

Besides being underused, China's vaccines—which contain inactivated virus—are less effective than the messenger RNA (mRNA) vaccines available elsewhere. A study in Hong Kong, where people could choose between the Pfizer-BioNTech mRNA vaccine or CoronaVac, a Chinese vaccine, showed three shots of either were more than 90% effective in preventing serious illness and death among those older than 60. But two doses of the mRNA vaccine were significantly more effective than two shots of CoronaVac. A separate study suggested CoronaVac's protection wanes more rapidly.

In May 2021, BioNTech and China's Shanghai Fosun Pharmaceutical agreed to set up a 50-50 joint venture to make and sell BioNTech's shot in China. But the product never received regulatory approval, perhaps to protect China's homegrown vaccines. Four Chinese companies have mRNA vaccines in various stages of development, but none has received the green light either. Pfizer's COVID-19 drug Paxlovid is available in China, as is a locally produced monoclonal antibody therapy, but it's unclear how many doses the country has.

Mainland authorities are still hoping to boost vaccination rates among the elderly. A 29 November directive of the National Health Commission urged local authorities to set up teams to go door to door and assess whether the ill and infirm can be safely vaccinated, and to launch education campaigns to counter vaccine hesitancy.

But those efforts will take time, says Xi Chen, a public health scientist at the Yale School of Public Health. "Reopening too soon will crowd out resources, crush the medical system, and cause more deaths," he says. For now, China remains trapped in the Catch-22 of its own making. ■

With reporting by Bian Huihui.



People in Beijing hold up empty sheets of paper to protest China's COVID-19 restrictions on 27 November.

PHOTO: THOMAS PETER/REUTERS



Scientist Suman Chandra checks cannabis plants at a growing facility at the University of Mississippi.

PHOTO: LANCE MURPHEY/*THE NEW YORK TIMES*/REDUX

U.S. SCIENCE POLICY

New U.S. law aims to light up medical research on cannabis

Biden signs bill streamlining pot studies and production

By **Meredith Wadman**

It just got easier for U.S. scientists to get their hands on some pot—for research, that is. President Joe Biden last week signed into law a bill that streamlines access to marijuana for medical research. The new law is expected to speed the issuance of government permits to scientists who want to study cannabis, whose medicinal promise has been widely touted but remains, with a few exceptions, unproved. It will also expedite permits for producers—including universities—that want to grow and distribute the drug for research. And the measure obliges the federal government to make sure an adequate, uninterrupted supply of marijuana is available to scientists.

"We will now be able to treat marijuana like we treat any other substance or pharmaceutical for which we hope there is potential benefit. We will be able to subject it to rigorous scientific trial," says Representative Andy Harris (R–MD), a physician and former National Institutes of Health (NIH)-funded researcher who helped usher the bipartisan legislation through Congress.

"This is exciting," says Ziva Cooper, director of the Center for Cannabis and Cannabinoids at the University of California (UC), Los Angeles. "The bill is a significant step forward with respect to chipping away at the barriers" to research. Scientists are eager to study cannabis and its derivatives as potential treatments for cancer, chronic pain, stress disorders, and other conditions.

Other cannabis researchers welcomed the law but said it doesn't go far enough. Some are disappointed it does not include a provision from an earlier draft of the legislation that would have allowed scientists to buy and study the marijuana available to consumers in the 37 states that have legalized its recreational or medical use, or both.

"There's no substitute for studying real-world products that our patients and recreational consumers are using," says Staci Gruber, a neuroscientist who runs trials of cannabis and its compounds as potential treatments for several conditions at McLean Hospital in Belmont, Massachusetts.

Cannabis is classified as a Schedule I drug—like heroin and LSD—under U.S. criminal law. That means both scientists and the growers who supply the drug for research must obtain permits from the Drug Enforcement Administration (DEA) and follow strict rules for storing and handling the drug. Researchers say it can take DEA a year or more to respond to permit applications.

The new law stipulates that within 60 days of receiving a researcher's application, the agency must approve it, request more information, or say why it is denying a permit. If a researcher submits more information upon DEA's request, the agency then has 30 days to make a decision. The law also makes it easier for researchers to amend their research protocols midstream, and ensures that DEA can't demand stricter security measures for marijuana than for other Schedule I drugs.

"The great advance of this bill is it puts a deadline on the [DEA's] response time," says Larry Walker, a pharmacologist at the University of Mississippi who has applied for research registrations.

Walker is the former director of a University of Mississippi center that, for decades, was the sole grower permitted by DEA to supply cannabis to U.S. scientists. But researchers complained that the center's pot was not nearly as potent as what is now commercially available. Since last year, DEA has registered six more growers, all of them companies, to produce marijuana for research.

Groff North America in Red Lion, Pennsylvania, is one of the newly licensed growers. Steven Groff, its founder and chief medical officer, dismisses complaints from scientists that DEA-registered companies can't provide cannabis products that mirror products widely available in legal dispensaries. "That myth is over," he says, noting his firm supplies smokable marijuana flower, vaping products, edibles, beverages, and soft chew tablets (gummies) to major universities. "We are going to be able to easily provide those materials now to researchers," he says.

Researchers can at least count on an ample supply going forward. The new law requires the attorney general to make an annual report to Congress on whether the DEA-controlled marijuana supply is uninterrupted and adequate for research needs, and to describe and remedy any shortfalls. And it orders NIH to produce a report within 1 year that addresses, among other things, how barriers to studying marijuana in states that have legalized pot might be overcome.

In the meantime, many researchers would like to see the federal government remove marijuana's Schedule I classification. That listing suggests cannabis is "uniquely harmful, and … medically unproven," even though U.S. regulators have approved a number of cannabis-based therapies, says Igor Grant, director of the Center for Medicinal Cannabis Research at UC San Diego.

In October, Biden ordered the attorney general to consider reclassifying cannabis. If it were moved down to Schedule II, pot would join a number of other approved drugs on that list, including morphine and Adderall, a treatment for attention deficit hyperactivity disorder. ■

FEATURES



Russian missiles targeted the Neutron Source facility (background) at the Kharkiv Institute of Physics and Technology.

PHOTO: NATIONAL ACADEMY OF SCIENCES OF UKRAINE

HERO CITY

Kharkiv was Ukraine's science hotbed until Russia attacked. The crippled city refuses to give up

By **Richard Stone**, *in Kharkiv, Ukraine*

In an SUV pockmarked by shrapnel, Mykola Shulga wends his way along the Kharkiv highway, dodging concrete barriers and antitank obstacles scattered along the road like giant toy jacks. On the northern outskirts of his broken city, he reaches the ruins of Pyatykhatky, an academic enclave set amid oak and maple groves. "This was one of the most beautiful areas of Kharkiv," he says, on a brisk October day. Before the war, it was home to many researchers at Ukraine's largest science center, the Kharkiv Institute of Physics and Technology (KIPT). But months of shelling left some of the district's apartment blocks uninhabitable. Shulga, a theoretical physicist and KIPT's director-general, gets quiet as the SUV passes a rubble pile where Pyatykhatky's hospital once stood. "They also destroyed the primary school," he says.

Russia's bloody war on Ukraine, now in its 10th month, has killed or injured at least 17,000 civilians, including scientists and students. It has also displaced more than 14 million, according to the United Nations. More than 1300 Ukrainian scientists—primarily women and men older than 60—have fled, finding refuge in labs in other countries. Tens of thousands of students are studying abroad.

Nowhere in Ukraine has science suffered more than here in Kharkiv, an academic community that Gerson Sher, co-chair of a U.S. National Academies of Sciences, Engineering, and Medicine (NASEM) working group on rebuilding science in Ukraine, calls "the crown jewel of Ukrainian science." In addition to KIPT, the Kharkiv region hosts nine other institutes of the National Academy of Sciences of Ukraine (NASU) and 65 universities and colleges. Few were left unscathed. Major losses include the control building for a unique NASU radio telescope and thousands of samples from a bombed national seed bank. The war has claimed 11 lives from V. N. Karazin Kharkiv National University, one of Ukraine's top academic centers, and a diminished student body has triggered mandatory layoffs of teaching staff.

Yet Kharkiv is resurgent. In a counteroffensive in September, Ukraine's army recaptured nearly the entire Kharkiv region—putting the city outside the range of Russian artillery fire for the first time in months. As winter draws near, academic centers here are buzzing with activity as staff repair blown-out windows. Some lab work and online classes have resumed, even amid blackouts and scant heating caused by Russian bombardment of the country's energy infrastructure. "Our enemy's plans to ruin our academic life were not realized," says Karazin rector Tetyana Kaganovska.

The resilience has energized efforts to staunch Ukraine's brain drain. Over the summer, the European Union launched a €25 million program, led by the Alexander von Humboldt Foundation, to place Ukrainian graduate students and postdocs in European labs with the expectation that they will return home when conditions allow. A host of smaller programs are helping scientists stay in Ukraine.

Alongside those emergency efforts, the European Union, the United States, and other allies have started discussions with Ukrainian counterparts on what a Marshall Plan for Ukrainian science might look like—if funding were to materialize. This month, the European Union is expected to approve a program in which DESY, Germany's largest accelerator center, will spearhead an effort to map Ukraine's science landscape, says DESY's Martin Sandhop. The exercise would inform any future makeover of Ukraine's many obsolete Soviet-era facilities and practices. "Ukraine needs to rebuild the system, not just the buildings," says Yury Gogotsi, a Ukraine-born materials chemist at Drexel University.

First, though, Ukraine must win the war—and many look to Kharkiv for inspiration. A billboard on the Kharkiv highway captures that sentiment: *Misto Heroi Kharkiv*, or "Hero City Kharkiv." That title recalls the honorific Soviet leaders bestowed on a dozen cities that withstood Nazi attacks during World War II, including four in Ukraine. Kharkiv did not make that list, but today it epitomizes Ukraine's resistance to Russian aggression.

Battle lines

Russian forces occupied Chornobyl and nearly seized Kharkiv soon after they invaded Ukraine on 24 February. Since then, the Ukrainian army has steadily clawed back territory. But Kharkiv, close to the border with Russia, remains vulnerable to missiles.



GRAPHIC: C. BICKEL/*SCIENCE*

"Our colleagues in Kharkiv sincerely believe in victory, that their city will stand despite all the troubles. These are not fake emotions," says NASU President Anatoly Zagorodny, who visited Kharkiv in August. But the city is down to half its prewar population of 1.4 million and is still in easy range of Russian missiles. Even Kharkiv's most ardent admirers acknowledge it will be a challenge to lure researchers and students back to rebuild the city's once-vibrant science scene.

IN THE FIRST DAYS of the war, Russia unleashed a blitzkrieg on KIPT, perhaps because of suspicions that the institute harbored a secret nuclear weapons program.

In a former incarnation a half-century ago, KIPT was in fact a weapons lab; it played a leading role in developing the Soviet bomb. The storied institute was founded in 1928 as the Ukrainian Institute of Physics and Technology by Abram Ioffe, widely regarded as the father of Soviet physics. Renowned theoretical physicist Lev Landau came a few years later, and the duo assembled a physics dream team, including the likes of George Gamow, a polymath who discovered the quantum tunneling mechanism for radioactive alpha decay and developed the big bang theory of our universe's origins. By 1946, Igor Kurchatov, architect of the Soviet nuclear weapons program, had anointed the institute "Laboratory No. 1" and closed its doors to the outside world.

After Ukraine became independent in 1992, KIPT emerged from the nuclear shadows as Ukraine's first National Research Center. It struggled to fund science amid the economic tumult and government corruption of the 1990s. But in a deal the U.S. government helped broker, from 2010 to 2012 KIPT and two other Ukrainian entities relinquished their weapons-grade, highly enriched uranium, shipping 234 kilograms of it to Russia, which blended it down to low-enriched uranium (LEU). In exchange, the U.S. Department of Energy promised to provide Ukraine with LEU fuel for power reactors, a program that KIPT now oversees. The United States also provided KIPT with funds and expertise to build a $90 million experimental reactor.



At the Institute for Single Crystals, a technician prepares samples in an ultra–low-humidity room.

PHOTO: RICHARD STONE

This facility, called the Neutron Source, would generate medical isotopes and beams of neutrons for probing materials. It would also serve as a prototype for subcritical nuclear power reactors that, unlike existing plants, do not sustain a chain reaction in their core. In August 2021, KIPT completed the reactor itself and a 100-megaelectronvolt accelerator that fires neutrons into the reactor to maintain fission reactions.

KIPT had just finished loading LEU fuel into the core and was ramping it up to a feeble initial power level of 20 kilowatts when war broke out. On 24 February, KIPT hurriedly shut down the core as explosions reverberated outside. In March and again in June, missiles and artillery shells damaged the building housing it and laid waste to its power supply, Shulga says. But the reactor vessel and its uranium fuel rods were unscathed. "That was a miracle," he says.

Elsewhere at KIPT, missiles blasted holes in buildings, blew out windows, gouged craters, and demolished the roof of its solid-state physics lab. After an International Atomic Energy Agency (IAEA) team visited last month, Director General Rafael Mariano Grossi called the damage "dramatic and shocking, even worse than expected."

The withering attacks hollowed out the institute: About half of its 2000 staff left Kharkiv, including dozens of elite scientists who found refuge in safer areas in Ukraine and 30 other countries. Seven are now at DESY. "Before the war, KIPT wasn't even on our radar screen," says DESY particle physicist Frank Lehner. DESY has since discovered KIPT's "extremely strong" work in theoretical physics, he says. KIPT's remaining staff are trying to sustain a minimal research program and repair salvageable buildings before winter.

CLOSER TO THE CITY CENTER, another NASU stronghold took heavy fire. Established in 1955, the Institute for Single Crystals (ISC) may be best known outside Ukraine for producing the lead-tungstate crystals and plastic scintillators used in particle detectors at the Large Hadron Collider in Switzerland (which discovered the Higgs boson) and the cesium iodide crystal detectors used in the Belle experiment in Japan (which observed asymmetries between matter and antimatter known as charge-parity violations). "Our materials helped win two Nobel Prizes," says ISC Director General Volodymyr Semynozhenko. ISC also exports industrial-grade sapphires and other specialty crystals, and it has diversified into chemical reagents and pharmaceuticals.

Until September, Valentin Chebanov, ISC's deputy director, camped out in his office, believing it was safer than his apartment downtown. The institute also converted a vast subbasement in its main building into a shelter with cots and a cooking area. Some 200 staff and their families holed up there for months. Bracing for a rough winter, ISC is installing internet and making other improvements to the shelter.

Shelling damaged ISC, although not nearly as extensively as KIPT. ISC staff

SCIENTISTS AT WAR

The volunteer

Olga Shpak made headlines around the world in 2017, when she captured the first footage of a pod of orcas using teamwork to hunt a young bowhead whale in the Sea of Okhotsk. The drone video was a sensational moment in her 2 decades of methodical fieldwork investigating bowheads, belugas, and other whale species—studies that had thrust the Ukrainian marine mammologist into the upper echelon of cetacean researchers and earned her a position at the A.N. Severtsov Institute of Ecology and Evolution in Moscow. "I loved my work. I could not imagine doing anything else," she says. Then the Russians stormed her homeland.

In February, as the war drums grew louder, Shpak took an overnight train from Moscow to Belgorod, near the Ukraine border, where her brother picked her up on the morning of the 23rd and drove her to their native Kharkiv. The very next day, the city was under siege. Shpak joined a local volunteer group and threw herself into sourcing medicine for civilians, and clothing and other supplies for the army. Thanks to her fluent English, she also serves as a liaison between Kharkiv's volunteers and international aid groups.

Several Russian colleagues emailed Shpak, expressing their shame over the war. "They cannot speak the truth in public," she says. She misses whales but has no time to think about science. "I don't know if I will ever work as a scientist again," she says. "My priority now is people. My fellow Ukrainians. This was a great discovery for me." —*R.S.*



After Russia invaded, Olga Shpak (second from right) set aside her research on whales to spend her days sourcing supplies for Ukrainian soldiers.

piled sandbags in front of windows in rooms with pricey instruments. But an electrical surge from the unstable grid fried the circuit board of a nuclear magnetic resonance machine, used for determining crystal structures. "You can imagine, no one wants to come to Kharkiv to fix it," Chebanov says. ISC researchers now must send samples to Kyiv—a 3- or 4-day turnaround. And they face chronic shortages of research-grade argon, helium, and other gases after a local supplier was bombed.

SOME 60 KILOMETERS southeast of Kharkiv, at the Semen Braude Radio Astronomy Observatory, a unique instrument is in shambles. The Ukrainian T-Shaped Radio Telescope, second modification (UTR-2) is the world's largest radio telescope tuned to decameter wavelengths, low-frequency emissions that hold clues to pulsars, the composition of the interstellar medium, and other phenomena. "They are pioneers of the lowest radio frequencies," says Philippe Zarka of the Paris Observatory.

Russian forces seized the Braude observatory on 25 February and used the 140-hectare grounds as a military base, says Vyacheslav Zakharenko, director of NASU's Institute of Radio Astronomy (IRA) in Kharkiv, which runs the observatory. After Ukrainian troops liberated it in early September, IRA staff found no damage to UTR-2's 2040 small dipole antennas, arranged in two arms extending 900 meters and 1.8 kilometers. But what they saw at the main building "was heartbreaking," Zakharenko says. Russian troops had looted or smashed instruments and computers and its roof was caved in. They'd also laid mines, including some in the service tunnels connecting the antennas to the central building.

Zakharenko's team has moved salvageable equipment into intact buildings before winter sets in. Replacing the electronics in UTR-2's control room "would be fast and cheap," says Michel Tagger, an astrophysicist at CNRS, the French national science agency, in Orleans. "That would make sense. It would keep the scientists there working," Zarka adds. But UTR-2 is 50 years old and becoming "more and more obsolete," he says.

IRA's long-term prospects may hinge on completing a successor to UTR-2 called the Giant Ukrainian Radio Telescope (GURT). It could detect distant exoplanets in a new way, by looking for the low-frequency radio flares whipped up when planetary magnetic fields interact with those of their host stars. By 24 February, IRA had installed five of GURT's 100 planned antenna arrays. When fully deployed, GURT would be 10 times more sensitive than UTR-2.

Tagger calls for a concerted European effort to help complete GURT and link it to the Low-Frequency Array (LOFAR), a radio telescope network in seven European countries that he and Zarka helped build. "You don't want to see 50 years of experience in radio astronomy in Ukraine wiped out," Zarka says. "Welcoming them into LOFAR would be a very good move."

FOUNDED IN 1804, Karazin is one of the oldest universities in Eastern Europe. Before the war, it had 23,000 students, including 4500 from overseas primarily studying medicine. When Kharkiv came under fire, about one-quarter of Karazin's facilities were destroyed, including schools for physics, business, economics, and public administration, as well as three dorms and a new sports complex. Nearly all the foreign students fled, as did many Ukrainian students. Scores of students and faculty joined the militia or volunteer units. Six teachers and five students are known to have perished on the battlefield or in shelling of civilian areas. "Every life lost is a profound grief for us," Kaganovska says.

PHOTO: OLGA SHPAK

Others stayed to protect the university. After a missile attack in March fractured a pipe and began to flood Karazin's archaeology museum, staff dodged artillery fire to come to campus and save collections from water damage. Faculty and students also rescued herbarium samples after the central heating system went down and salvaged instruments from the School of Physics and Technology in Pyatykhatky.

Karazin's classes went fully online in March. It has also struck deals with universities in western Ukraine and in other European nations to host hundreds of its students as they earn Karazin degrees. That has helped Karazin hold onto about 85% of its Ukrainian students. "Preserving their education is the most important way we can help our country," Kaganovska says.

But the university faces financial woes. Because the central government pegs university budgets to the number of enrolled Ukrainian students, it reduced Karazin's funding for the current academic year by 15%. Karazin also lost tuition revenue from foreign students. Forced to make painful cuts, Kaganovska, a law professor, did not renew contracts for 48 faculty members in the School of Physics, including its chair. Other departments fared better, which "caused enormous outrage," says Eugene Chudnovsky, a physicist at the City University of New York who grew up in Kharkiv and earned his Ph.D. from Karazin. The university, he says, "has always been a major supplier of scientists to Kharkiv's research labs."

Kaganovska says the student-teacher ratio has been unusually low in Karazin's physical sciences departments, compelling the university to tap reserves to cover salaries—money the war has largely drained. She reinstated the laid-off physicists last month, after a U.S. donor stepped up to cover part of the shortfall. "We are looking for the rest now," Kaganovska says. "Definitely our mission is preserving science."

When peace arrives, Karazin is counting on far-flung faculty and students to return to campus. Several universities here, including the University of Biotechnology, have pulled up stakes and now operate in western Ukraine. Kharkiv can't afford further losses, Chudnovsky says. "The future of Kharkiv will very much depend on education in Kharkiv."

AS THE WAR GRINDS ON, scientists across Ukraine are looking beyond a harsh winter of power cuts to an expected victory—and the prospect of rebuilding tattered institutes and universities.

An overriding priority is preventing a further loss of talent. A mad scramble in the spring to place refugee-scientists in labs outside Ukraine has now evolved into a widespread effort to help more than 50,000 scientists who have stayed. Many are men under the age of 60 who are forbidden from crossing the border without explicit permission, in case they are needed for the military. (For now, scientists and students are exempt from service.) Thousands of women scientists, too, have stayed to care for family members or serve in the military or as volunteers.

At least three dozen assistance programs in Europe and the United States are now throwing lifelines to scientists in Ukraine. Many are small efforts, targeted to specific disciplines. For instance, an initiative of the Institute of Electrical and Electronics Engineers, a large U.S. society, funded 22 Ukrainian proposals on magnetism to the tune of $180,000 this year and plans to spend another $100,000 on new projects in 2023.

But wartime conditions hamper research. "You can't do science without money. Without electricity," says Jerzy Duszyński, president of the Polish Academy of Sciences (PAN). So NASEM and PAN are testing an innovative strategy. With $3 million from the Breakthrough Prize Foundation, they launched a program in October that will allow up to 30 groups led by elite scientists in Ukraine to spend up to 3 months a year doing hands-on research in the labs of Polish collaborators—and pay their salaries for the remainder of the year

SCIENTISTS AT WAR

The caretaker

Yuliia Guglya says it was always a dream to live where she works. That dream has now come true—but as part of a nightmare. In her work as an entomologist, Guglya had spent countless hours in the collections of the Museum of Nature at V. N. Karazin Kharkiv National University. But when Russia invaded Ukraine on 24 February, she became its resident curator. She, her husband, and their 21-year-old son moved into a warren of rooms in the museum's basement, where she has stashed some of the 215-year-old collection's prize holdings.

Guglya's first love was butterflies, before her graduate adviser implored her to switch to leaf miner flies, which no one at the time was studying in Ukraine, she says. "I was so sad. I thought flies were disgusting." Guglya came around to them and has discovered 25 species in the leaf miner family. Many more remain to be described among her 40,000 specimens, she says. "I couldn't evacuate my collection. And anyway, I didn't want to escape." For weeks, Guglya, museum director Rostyslav Luniachek, and a few colleagues worked to move the most coveted specimens to the basement. They then turned to boarding up the 200 or so windows blown out by bomb blasts. "We're lucky," she says. "My husband's a carpenter."

Guglya doesn't know when she and her family—her parents-in-law joined them in April—will vacate the museum. Perhaps sensing her distress, a young black cat, named Javelin—after the U.S. shoulder-fired antitank rocket—ambles over to her makeshift desk and plops onto her lap. "He helps us to be here," she says, stroking Javelin's fur. But she does not regret her choice. "We are saving the museum. And the museum is protecting us." —*R.S.*



Yuliia Guglya moved her specimens to safety in the Museum of Nature's basement.

PHOTO: RICHARD STONE



Dipole antennas at a pioneering radio telescope outside Kharkiv, Ukraine, were not damaged during Russia's monthslong occupation.

in Ukraine. Such a lifeline, Duszyński says, "will help keep Ukrainian science alive." Over time, Duszyński hopes the elite teams will nucleate Ukraine's scientific revival, by strengthening their connections to top institutions all over the world.

For KIPT, German institutes are emerging as potential saviors. DESY aims to bring KIPT scientists into a new quantum electrodynamics experiment at the European X-ray Free-Electron Laser, a facility that has suspended collaborations with Russia. And the GSI Helmholtz Centre for Heavy Ion Research in Darmstadt, Germany, is looking to involve KIPT in FAIR, a particle accelerator under construction that would use high-intensity beams to create matter that now only exists in the universe's harshest environments, such as supernovae.

Embedding Ukrainian scientists in Europe's premier science facilities makes sense as Ukraine grapples with what research infrastructure to rebuild—and where—after the war, says theoretical physicist Yulia Bezvershenko, a visiting scholar at Stanford University. "We should build something they don't have" in Europe.

She and others hope peace will open the way to an overhaul of Ukraine's R&D enterprise. The government's own draft recovery plan for science and education calls for setting up a genomics research center and a Ukrainian version of the U.S. Defense Advanced Research Projects Agency, which funds high-risk research. The plan also calls for rebuilding Ukraine's small oceanographic research fleet.

Another objective is expanding merit-based funding. The main source of competitive grants is the nascent National Research Foundation of Ukraine (NRFU), which in 2020 allotted $9 million to 216 projects. NRFU's $25 million budget for this year was sequestered for the military. But in 2023 it plans to spend $12 million on 57 projects related to security and sustainable development, says NRFU Executive Director Olga Polotska.

Gogotsi gives NRFU high marks for recruiting expats and other scientists outside Ukraine as reviewers. He's less flattering about other parts of Ukraine's research establishment—especially NASU, which he says is largely unchanged since the Soviet era. Any infusion of postwar funding for Ukrainian science, he argues, should be tied to integrating NASU's many institutes into the university system and getting them to compete for funding. He also wants to see them become "greenhouses" for tech startups.



Russian troops looted and smashed instruments in a radio telescope's central building.

Although NASU's Zagorodny recognizes the need for reform, he says he wants to "maintain our historical traditions." He rejects the idea of subordinating NASU institutes to universities. As a better model, he points to Germany's Max Planck Society's world-class institutes, which thrive independently alongside the country's university system.

Like many in Ukraine, Shulga is not waiting for peace before planning a revival. In October, at an IAEA conference in Washington, D.C., he met with Grossi, who he says supports KIPT's aim to establish an International Center on Nuclear Physics and Medicine with the Neutron Source facility as its centerpiece.

Bezvershenko, however, says it's risky to build major scientific infrastructure near the Russian border. Chudnovsky has even floated the idea of relocating KIPT, if only temporarily, to western Ukraine. That move would echo the Soviet response to the Nazi invasion of Ukraine during World War II, when it evacuated Kharkiv's science establishment to eastern Russia.

But Shulga has no plans to move, not least because it would be impossible to uproot the Neutron Source and other large KIPT facilities. More importantly, such a retreat would undermine the ambitions of Ukrainian science leaders to restore the luster of Kharkiv, the nation's hero city and its scientific jewel, Zagorodny says. "That's how we will honor its courage and heroism." ■

Richard Stone's Ukraine reporting was supported by the Richard Lounsbery Foundation.

PHOTOS: (TOP TO BOTTOM) OBSERVATOIRE RADIOASTRONOMIQUE DE NANÇAY; ANATOLY MYASOED/INSTITUTE OF RADIO ASTRONOMY



OUT OF THE ASHES

Slavutych was built after the nuclear meltdown at Chornobyl. The town, scarred by war, is planning a research revival

By **Richard Stone**, *in Slavutych, Ukraine*

In one of the last acts of camaraderie in a fracturing nation, Soviet laborers from the Baltics to the Caucasus converged on a pine forest in the late 1980s to build a Ukrainian town from scratch. Slavutych was a new home for workers from the Chornobyl Nuclear Power Plant and their families after its 1986 explosion turned Pripyat, the town next to the plant, into a radioactive wasteland. In March, Slavutych's 20,000 residents endured a different breed of terror, as Russian troops massed outside the town.

Heroic dashes around enemy lines to secure supplies—and a potent display of solidarity in the town's central square—staved off disaster. Now, Slavutych, like the rest of Ukraine, is girding for a grueling winter of electricity outages. But Mayor Yuri Fomichev and Anatolii Nosovskyi, a radiation warrior with decades of experience at Chornobyl, are already planning Slavutych's postwar rebirth as a science center, one that would focus on a formidable challenge: dismantling the radioactive remains of Chornobyl's destroyed Unit 4 reactor.

Lessons learned at Chornobyl—the mother of all decommissioning projects—can be applied globally, at dozens of nuclear plants that are slated to shut down in the coming years, says Nosovskyi, director of Ukraine's Institute for Safety Problems of Nuclear Power Plants (ISPNPP) in Kyiv. "Our idea is to bring together specialists from across Ukraine in Slavutych. It's what the town was always meant to be—a research cluster."

Creating a science center "in an environment that's unique in the world would very much make sense," says Kai Vetter, a nuclear physicist at University of California, Berkeley, whose team has donated instruments and supplies to ISPNPP since the war started. "It's a fantastic idea," adds Nick Tomkinson, a nonproliferation expert at Global Nuclear Security Partners, a London consulting firm that hopes to map radioactive contamination around Chornobyl.

On 24 February, Russian troops swarmed across the border with Belarus and seized the Chornobyl plant. Nosovskyi, who began his career working on nuclear submarine radiation safety, worried the invaders would reawaken the radioactive nightmare he witnessed when he was dispatched to Chornobyl in 1987. There, he monitored the radiation received by tens of thousands of scientists and soldiers as they erected a concrete shelter over Unit 4's seething remains. Over the years, he worked hand in hand with Russian scientists—cooperation that ceased after Russia's illegal annexation of Crimea in 2014. "We figured out how to manage the Chornobyl problem on our own," he says.

In April, after Russian soldiers retreated from the region, Nosovskyi ventured back to a satellite ISPNPP facility in Chornobyl. Smashed instruments and glass shards carpeted its chemistry lab's floor. Looters had made off with a dozen vehicles and the facility's newest computers. Scores more computers had been stripped of hard drives. "I gave up smoking 4 years ago. Took it up

PHOTO: RICHARD STONE


A control room in Chornobyl's Unit 4 reactor has been empty since 1986, but the reactor's remains still seethe.

again after seeing what happened to the lab," Nosovskyi says, lighting up a cigarette. (As he talked to *Science* at an outdoor café on 10 October, three Russian cruise missiles whizzed by overhead on their way to targets in western Ukraine.)

Nosovskyi believes Russia raided the Chornobyl lab for evidence that Ukraine was working to secretly develop nuclear weapons, as three Russian newspapers falsely alleged in late February. "Of course, they made that up," he says, pointing out that Ukraine's nuclear facilities are monitored by the International Atomic Energy Agency.

Russia's military also took plant workers and guards hostage. A chilling moment came when two officers demanded access to one of Chornobyl's two spent fuel repositories, which harbor ferociously radioactive material. The crew chief rebuffed them, says Valeriy Seyda, Chornobyl's acting director, and the suspicious pair left without incident.

But the occupiers stole equipment from the plant and tracked in radioactive contamination after digging trenches and laying mines in the nearby Red Forest—named for pines killed by a plume of radiation from the 1986 accident. Since the soldiers' retreat, the plant complex has been cleaned up and parts of Chornobyl village have been demined, but it remains a headache to transport staff from Slavutych to Chornobyl for decommissioning tasks at the plant's four shuttered reactors and maintaining the spent fuel repositories. Before the war, workers could make the 50-kilometer commute by train through Belarus. With that border closed, staff endure a 340-kilometer, 6-hour bus ride around Belarus, and work 8-day shifts requiring them to bunk in Chornobyl.

"It's what the town was always meant to be—a research cluster."

Anatolii Nosovskyi, Institute for Safety Problems of Nuclear Power Plants

Slavutych, meanwhile, was devastated economically. In the months before the invasion, the town had become a base camp for foreign tourists inspired to visit by the HBO miniseries *Chernobyl*. The war crushed that cottage industry. In late February, a bridge on the only road into Slavutych was blown up, and a Russian battalion camped outside the town. As supplies diminished, volunteers undertook perilous drives on logging roads—dodging shells and Russian patrols—to reach villages outside the blockade and spirit back staples such as milk, potatoes, and flour. "That allowed us to survive," Fomichev says.

In early March, natural gas supplies to Slavutych were interrupted, shutting down the town's communal heating plant. Engineers rigged a boiler to run on firewood, which managed to supply just enough heat to keep apartment blocks from freezing. Then Russian troops sabotaged the main powerline to Slavutych. The lights went out—as did electric stoves. Residents resorted to cooking over open fires on the streets in frigid temperatures.

In the meantime, Slavutych organized a militia: about 200 volunteers, a handful of whom had military training. Equipped only with Kalashnikov rifles, they engaged Russian tanks and artillery on the town's outskirts. "We put up a brave fight," Fomichev says. Five died in skirmishes, and badly outgunned, the militia on 25 March agreed to a Russian demand to leave the city. But when a Russian convoy entered the town square, it encountered a throng of unarmed protesters chanting "Slavutych is Ukrainian, go home!" "The commander had no idea what to do. Apparently, he didn't have the stomach to slaughter unarmed civilians," Fomichev says. Two days later, the battalion withdrew.

The town's struggles are not over. Over the past several weeks, Russian bombardments have left Slavutych without electricity most hours of the day, and town officials worry about further disruptions to the supply of natural gas.

But Fomichev and Nosovskyi are already proselytizing for their vision of Slavutych as a nuclear science hub specializing in radiation medicine, radioecology, and the monumental task of decommissioning the Unit 4 reactor, an effort expected to last at least 40 years. Research is sorely needed in such areas as radiation-hardened robotics and the properties of irradiated graphite from the destroyed reactor's core. "We really don't know how to safely handle such material," Nosovskyi says. "There are enormous opportunities to develop and demonstrate advanced technologies," Vetter says.

Nosovskyi dreams of eventually installing a small modular reactor, built in a factory elsewhere and shipped to Slavutych to generate power and to use as a training facility. He envisions launching a training program on reactor decommissioning with the help of the Igor Sikorsky Kyiv Polytechnic Institute, which already has a branch in the town. With the train link down for the foreseeable future, Fomichev is exploring a ferry service on the Dnieper River that would speed the commute between Slavutych and Chornobyl.

Realizing the vision would require government or international funding, a possibility only after the war ends, Nosovskyi acknowledges. But the war has strengthened his resolve. He mentions the grandson his daughter gave birth to in Kyiv in March while holed up in an underground shelter. "The little boy doesn't know yet how brave he was. How brave his mother was. But she named him Lev," the Ukrainian word for lion. "So he will know, and remember, someday." ■



ILLUSTRATION: CHIARA VERCESI

AAAS/SUBARU SB&F PRIZES FOR EXCELLENCE IN SCIENCE BOOKS

New books for young scientists

A charming science-inspired spell book, an overdue ode to overlooked scientists, a vivid exploration of the circle of life, and a dazzling tome packed with images of the cosmos are featured among the finalists for the 2023 AAAS/Subaru SB&F Prize for Excellence in Science Books, a collection of awards sponsored by Subaru of America, Inc., in partnership with the American Association for the Advancement of Science (AAAS, the publisher of *Science*). Read on for reviews of the finalists—which take on topics ranging from the fungal to the forensic—written by *Science* staff with the help of a few friends. **—Valerie Thompson**

CHILDREN'S SCIENCE PICTURE BOOK

Fox

Reviewed by **Sacha Vignieri**[1]

Understanding death is a challenge for most, but perhaps especially so for young children. Yet sometimes even they must grapple with this phase of the life cycle. In *Fox: A Circle of Life Story*, Isabel Thomas tells a story about the life and death of a mother fox. The book begins by playfully describing the mother fox's life as she hunts and cares for her kits as they grow, play, and learn. Midway through, the fox meets an untimely end when she is hit by a car after a successful hunt. Her body, readers learn, provides food for scavengers and decomposers, and eventually all that was once the fox is recycled into new life. At the end of the book, Thomas provides a more detailed description of the decomposition and recycling process, describing, in an accessible way, the science behind the book's narrative. The story, warmly illustrated by Daniel Egnéus, not only teaches the biology of death, decomposition, and renewal but may provide solace to children affected by death or dying.

And yet, while the story is sweet and the illustrations adorably depict the fox and her family engaging in typical fox behaviors in the book's first half, the mother's death feels jarring, especially because much has been made of fox family's bonds before this event. A child might wonder, "What will happen to the fox kits? How will they survive?" The book's illustrations suggest that they do survive, but in nature, they probably would not have made it through the winter alone. For most children, this book's theme might be too upsetting, but for those who have lost someone they love, it may help them to better understand the ways death makes way for new life.

Fox: A Circle of Life Story, *Isabel Thomas, Illustrated by Daniel Egnéus*, Bloomsbury Children's Books, 2021, 48 pp.

Good Eating

Reviewed by **Jennifer Sills**[2]

In *Good Eating: The Short Life of Krill*, author Matt Lilley challenges readers to look at the world through the eyes of a krill, a tiny crustacean that serves as a food source for many marine animals. "You" (the reader, embodying the krill's experience) sink down into the dark ocean as an egg, hatch, and swim back to the surface, molting repeatedly. Lilley describes each stage of the krill's life cycle in simple language. When the egg first hatches, he explains, "You are a six-armed oval." Later: "You grow more spines." Just



before the krill reaches the surface, he quips, “You grow a face…sort of. About time!”

As the krill matures, the reader discovers its special skills: Your transparent stomach can turn green! You can light up! Now at the water’s surface, you learn that you are not alone: There might be a “million million” creatures just like you—perhaps “a krillion”? (The appendix, which provides background about the Antarctic krill, clarifies that a million million is, in fact, a trillion.)

Dan Tavis illustrates the krill’s journey with clean ink lines and watercolor shading. The krill’s red hues stand out against a backdrop of cool blues and greens. The images reinforce the reader’s role by positioning predators as a krill might see them. The hungry faces of penguins, slightly distorted as if viewed from underwater, peer down from an opening in an ice sheet. The mouth of a blue whale fills a subsequent two-page spread. The book’s main character pops with personality, thanks to its googly eyes and entertaining facial expressions. Although the text plainly describes the krill’s most important role—eating single-celled organisms until becoming food itself—there is no need to worry. The little krill around whom the story revolves escapes its predators and lives to continue its adventure…this time.

Good Eating: The Short Life of Krill,
Matt Lilley, Illustrated by Dan Tavis, Tilbury House Publishers, 2022, 36 pp.

HeroRat!

Reviewed by **Jeffrey Mervis**[3]

HeroRat! tells the story of Magawa, an African giant pouched rat who has been trained to sniff out buried land mines. The book’s 32 pages are packed with colorful and action-filled drawings.

There is also plenty of content. *HeroRat!* is stuffed with details about Magawa’s training, including the three reasons that pouched rats “make better heroes” than dogs (they cost less to keep and train, they can easily work with different handlers, and they do not weigh enough to trigger mines). The main character has a can-do personality bordering on braggadocio. “I thought it was going to be hard to become a hero, but I’m really good at searching for those awful explosives,” Magawa says about his training by a Tanzanian nonprofit before heading off to Cambodia, which we learn ranks second to Afghanistan as the country with the most land mines. There, he performs so well that he garners a gold medal “for animal gallantry,” billed as the first ever awarded to a rat.

The book’s somber message—that animals can help the planet recover from the ravages of war—is probably too complicated, if not downright scary, for little children. But having Magawa deliver a first-person account of his training renders it a bit too childish for older students. So the book’s target audience is not obvious. Asked what he liked best about *HeroRat!*, my 9-year-old grandson replied, “You normally think of rats as bad. But this book shows how they can help people.” Adults may want to consider adding this simple but important message as they share Magawa’s story with young listeners.

HeroRat!: Magawa, a Lifesaving Rodent,
Jodie Parachini, Illustrated by Keiron Ward and Jason Dewhirst, Albert Whitman & Company, 2022, 32 pp.

Tu Youyou’s Discovery

Reviewed by **Caroline Ash**[4]

In 2015, Tu Youyou became the first Chinese woman to win a Nobel Prize, sharing the Prize in Physiology or Medicine for her role in identifying the ingredient of an herbal remedy that is now the world’s most important antimalarial drug. *Tu Youyou’s Discovery*, by Songju Ma Daemicke, tells the scientist’s inspiring story.

At a time when most girls in China were not formally educated, Youyou’s parents sent her to school, but at 15 years old, she contracted tuberculosis and was bedridden for several months. With the help of antibiotics and her mother’s traditional remedies, Youyou made a full recovery over the next two years. This experience became her inspiration, and she persisted with her education to become a research scientist in 1955, applying modern scientific methods to

test traditional medicines as potential treatments for drug-resistant malaria.

With clear text and vivid illustrations that will appeal to a range of age groups, the book describes how Youyou, whose work overlapped with China's Cultural Revolution, overcame numerous challenges, from limited resources to doubts sewn by her male colleagues. It also reveals how an ancient text inspired the simple method that would ultimately be used to extract artemisinin, or qinghaosu—the compound that would prove effective against malaria—from sweet wormwood. By the time Youyou stood on the podium in Stockholm to receive her prize, artemisinin had saved the lives of 6.8 million children.

The book may be slim, but it does an excellent job of conveying a sense of the time, sacrifice, and the sheer persistence that often underlie revolutionary scientific discoveries. It also succeeds in showing that the scientific method can be applied to folk knowledge to great effect and in celebrating an otherwise unsung life in science.

Tu Youyou's Discovery: Finding a Cure for Malaria, *Songju Ma Daemicke, Illustrated by Lin, Albert Whitman & Company*, 2021, 32 pp.

MIDDLE GRADES SCIENCE BOOK

It Takes Guts

Reviewed by **Marc S. Lavine**[5]

The journey our food takes from the time it enters our bodies to when it exits is long and complicated, and it involves many muscles, tissues, and organs. Capturing the details of each step along the way could be an arduous task, but Jennifer Gardy manages to describe the processes of human digestion in a lively way in *It Takes Guts*.

Each chapter of the book covers a part of the digestive tract, from the mouth to the stomach to the intestines, and describes the different stages of digestion that occur in each one. Readers learn how two sphincters in the esophagus control the passage of food downward while preventing the upward passage of stomach acid, all under the automated control of our enteric nervous system. Sword swallowers train themselves to actively control these sphincters to create an almost straight tube from mouth to stomach, a skill exploited by the 19th-century German physician Adolf Kussmaul, who attempted to map out the esophagus of one such individual using a long tube lined with mirrors.

Further along, readers learn about mucus and enzymes, gastrin and ghrelin, and villi and crypts, all of which play roles in regulating eating and digestion. Less directly involved, but just as important, are other organs that participate in the digestive process. The liver produces bile acids, the pancreas produces insulin and glucagon, and even the appendix plays a part—it is now believed to act as a storage chamber for beneficial bacteria. Speaking of bacteria, many are essential to proper digestion, as some produce enzymes that humans need but cannot synthesize.

Gardy also includes a discussion of normal and not-so-normal things everyone experiences from time to time—such as burping, farting, and more. Being able to distinguish healthy from unhealthy functions is often key to diagnosing ailments and disease, so this material may be particularly valuable to the reader. Overall, *It Takes Guts* is an interesting, entertaining, and informative book to bite into.

It Takes Guts: How Your Body Turns Food into Fuel (and Poop), *Jennifer Gardy, Illustrated by Belle Wuthrich*, Greystone Kids, 2021, 152 pp.



ILLUSTRATION: BELLE WUTHRICH

Snoozefest

Reviewed by **Seth Thomas Scanlon**[6]

Fewer and fewer adolescents are getting enough shut-eye (*1*, *2*). *Snoozefest* is a delightful and informative new book on the science of sleep that may just convince budding young night owls to catch a few more z's every evening.

Author Tanya Lloyd Kyi delves into the history of sleep research with bite-sized but detailed summaries of important experimental milestones, such as the development of electroencephalography (EEG), the characterization of melatonin, and the discovery of circadian rhythms. She also highlights the latest findings on why we sleep and dream, noting, for example, how it helps with consolidation of memories, formation of insights, recovery from trauma, integration of emotions, and repair of the brain's daily wear and tear.

The book describes how sleep can be beneficial for the proper functioning of the metabolic, cardiovascular, and immune systems, while acknowledging that many major questions remain unanswered. Along the way, Kyi touches on some important public policy issues such as the dangers of sleeping pills, the problems associated with night-shift work, and the need for school systems to better accommodate the "phase-delayed" sleeping patterns of most teens. The author even provides some fascinating examples of how various other species have managed to balance the risks of predation with the need for sleep.

A few tangential topics, such as Freudian dream analysis, are a distraction, and the attention paid to alternative "treatments" for insomnia (e.g., acupuncture) is, in my view, problematic. Instead, the author might have devoted more space to subjects such as how animals dream and sleep. The drawings by Valéry Goulet are fun and inviting but mostly ornamental—the relative dearth of scientifically informative illustrations feels like a missed opportunity.

Still, this book is certainly not the "snoozefest" suggested by the title. Rather, it is a very engaging and informative volume on an enormously important area of research that should rouse even the most bleary-eyed teen from their slumber.

Snoozefest: The Surprising Science of Sleep, *Tanya Lloyd Kyi, Illustrated by Valéry Goulet*, Kids Can Press, 2021, 80 pp.

REFERENCES AND NOTES

1. J. M. Twenge, Z. Krizan, G. Hisler, *Sleep Med.* **39**, 47 (2017).
2. S. Paruthi *et al.*, *J. Clin. Sleep Med.* **12**, 1549 (2016).

Stolen Science

Reviewed by **Yevgeniya Nusinovich**[7]

When children learn about science, they are often given the impression that most major discoveries were made by white men. However, *Stolen Science* reveals that there is often more to the stories that perpetuate this idea, and for every Marie Curie who was given credit for her work, there are many women and others throughout history whose scientific contributions went unacknowledged.

In this book, Ella Schwartz presents the stories of 13 scientists and inventors who contributed much to science and in many cases are still forgotten. A pattern emerges across each story: A woman, an enslaved person, or another member of a marginalized group makes a major discovery or invention, but a white man or men get all the credit. In some cases, the men in question intention-



ILLUSTRATION: GABY D'ALESSANDRO

ally undermined the person who made the discovery, as happened to crystallographer Rosalind Franklin, whose crucial work on the structure of DNA was minimized by her competitors. In other cases, the individual's contribution was recognized only after it was later confirmed by someone else, as happened to Carlos Juan Finlay, a Cuban physician who identified how yellow fever is transmitted. The world gave credit to US army physician Walter Reed, even though Reed himself readily acknowledged that Finlay had made the discovery.

Each story in this book is supported by a separate section of text explaining the science behind the work in question using terms that middle-school-aged or even younger children should be able to follow. In addition, each of the featured individuals is beautifully portrayed by illustrator Gaby D'Alessandro, who combines images of the scientists and inventors with symbols of their work.

This is a very important book that not only gives credit where it is long overdue but should also help inspire future scientists from all backgrounds. Readers will be left wondering how much better the world would be if everyone was allowed to develop their talents to the fullest.

Stolen Science: Thirteen Untold Stories of Scientists and Inventors Almost Written out of History, *Ella Schwartz, Illustrated by Gaby D'Alessandro*, Bloomsbury Children's Books, 2021, 128 pp.

How to Build a Human

Reviewed by **Sacha Vignieri**[1]

We humans attribute our domination of Earth to our big brains, which have allowed us to adapt to, and convert, landscapes all over the planet. However, our oversized cortex is a relatively recent development. In the introduction of her book *How to Build a Human*, Pamela Turner emphasizes that just 3 million years ago, an extraterrestrial looking for the animal with the largest brain would have selected an ancestor of modern dolphins.

In the pages that follow, Turner takes the reader step by step through human evolution, from an ancient ape ancestor that lived around 20 million years ago to today. Our understanding of this lineage has greatly expanded over the past few decades, and she cleverly describes how humanity's evolutionary tree is full of many branches, each distinctive but in many ways similar. The journey takes place in seven steps, beginning with "We stand up," which describes the evolution of bipedal locomotion, an innovation that likely first emerged in African apes around 7 million years ago. It concludes with "We become storytellers," which summarizes how early *Homo* species developed art, a likely expression of more complex language skills.

Engaging illustrations by John Gurche, a respected paleoartist, accompany Turner's story. His artwork conveys the myriad species of hominoid in ways that clearly display their "humanity," rendering them in ways that allow readers to recognize these species as our relatives. The book is also full of photographs of fossils, art, and tools that further bring human ancestors and relatives to life. Readers will leave this book with an understanding of how we came to be who we are and with an appreciation for the fact that, for much of our history, we lived alongside other species descended from the same ancient ancestor.

How to Build a Human: In Seven Evolutionary Steps, *Pamela S. Turner, Illustrated by John Gurche*, Charlesbridge, 2022, 176 pp.

HANDS-ON SCIENCE BOOK

Forensics for Kids

Reviewed by **Michael A. Funk**[8]

As we move through the world, we leave traces of ourselves that can lead back to us. When a violent crime occurs, police can turn to field and laboratory analysis to determine what has happened at the crime scene and whether a suspect was present. In *Forensics for Kids*, Melissa Ross traces how forensic science emerged in Britain and the United States in the 1800s as policing practices mixed with rudimentary medicine and chemistry. This combina-

tion has been refined over the past two centuries, with new sources of evidence, analytical methods, and standards required for achieving a guilty verdict.

Forensics for Kids consists of briefly described anecdotes and overviews of specific cases, many of which will be familiar from detective novels and news stories. The scientific explanations of how various forensic techniques are or were applied are short, but the book does include many activities that children can do to simulate the process of collecting and analyzing evidence. For example, readers can try fingerprint matching, identifying a mystery kitchen chemical, or creating a database of shoe prints. Many of these activities can be done in a group and might inspire games or creative adaptations.

Although detective work is important to solving crimes, the book does not examine how systemic biases or unreliable, unscientific forensic methods might distort criminal trials and lead to wrongful convictions. Blood splatter and bite mark analyses, for example, are presented with no mention that the accuracy of these methods has been vigorously questioned by the scientific community.

In one section, the book describes automated DNA testing and aggregated DNA databases as tools to "keep dangerous criminals from being released to the streets," without discussion of the privacy and due process concerns these methods can pose. Although this context may seem unnecessary, it is critical for readers to understand that such methods have limits and that errors can result in serious consequences for defendants.

Forensics for Kids: The Science and History of Crime Solving, With 21 Activities, *Melissa Ross*, Chicago Review Press, 2022, 144 pp.



ILLUSTRATION: KELLY ANNE DALTON

Physics for Kids

Reviewed by **Pamela J. Hines**[9]

With *Physics for Kids*, Liz Lee Heinecke adds to her series of books dedicated to hands-on, at-home science experiments. Each chapter of this new installment begins with a full-page color portrait of a scientist whose discoveries underlie the chapter's experiments. The individuals chosen represent the diversity that enriches scientific discovery. A short biography introduces the reader to the featured scientist and describes how a particular discovery they made is applied in the real world. The book then provides instructions for hands-on experiments that explore topics such as relativity, thin films, and sunspots, to name just a few.

Readers learn about Italian physicist Laura Bassi, born in 1711, who studied atmospheric electricity at her country house when experiments on lightning were banned in the nearby city of Bologna. Heinecke suggests exploring Bassi's fascination with static electricity by watching how an electrically charged balloon can chase soap bubbles on a plate. Another researcher, American engineer Christine Darden, born in 1942, worked at NASA doing calculations for the first space missions and went on to study aircraft wing design. Heinecke encourages readers to follow in Darden's footsteps by folding paper airplanes to see how wing shape affects aerodynamics.

The experiments in this book are well chosen to suit students in the upper grades of elementary school or in middle school, and the supplies needed are simple and readily available. Safety tips are established early on, keeping the focus on carefully exploring the natural world. For each set of experiments, which are illustrated by photographs alongside the how-to protocol, a short paragraph explains the physics that drives the phenomenon under consideration. For those who want to dig deeper, the author suggests potential modifications. This would be a handy book to have around for any youngster interested in the physics of the natural world.

Physics for Kids: Science Experiments and Activities Inspired by Awesome Physicists, Past and Present, *Liz Lee Heinecke, Illustrated by Kelly Anne Dalton*, Quarry Books, 2022, 128 pp.

The Science Spell Book

Reviewed by **Valerie Thompson**[10]

As alchemists and astrologers were supplanted by chemists and astronomers, many mysterious elements of the natural world yielded to scientific explanations. But this does not mean that there is no longer a place for magic in science. "Creativity and awe are integral traits of any success-

ful scientist, and they are rooted in the imaginative exercises of childhood," notes Cara Florance in *The Science Spell Book*, a compendium of mystical experiments for the science-minded spell-caster.

The collection's 25 activities are divided into five categories: "infusions," exploring plants as pigments and indicators; "illumination," probing the nature of light; "sorcery," interrogating electromagnetism and other fundamental forces; "alchemy," describing physical and chemical changes; and "mimicry," looking to plants and animals for engineering inspiration. Moody photos featuring minerals, dried herbs, and candles illustrate the activity's end products—a violet solution that changes color when breathed upon; earthy pigments derived from coffee, flower petals, and clay; insect wings and acorns coated in borax crystals.

Fanciful titles and charming incantations infuse each activity with magical touches, but science lurks just below the surface. "Current flows through the coiled course. Domains align to exert their force," reads the opening of "Magnetic Manifestation," for example, which guides readers through the steps required to create an electromagnet. A short summary of the principle at play comes next, followed by a list of supplies, steps, and a short section that elaborates on the science underlying the activity. Readers are likely to have some of the required items on hand—rubbing alcohol, cream of tartar, and turmeric, for example. Others are more unusual—think butterfly pea flower tea, hobby motors, LEDs—and might need to be purchased in advance.

"Learning the scientific reasoning behind a mysterious event does not take away from the magic but instead exposes more exciting mysteries and questions to be answered," observes Florance. Readers need look no further than this delightful book for proof that this is so.

The Science Spell Book: Magical Experiments for Kids, *Cara Florance*, Sourcebooks eXplore, 2022, 224 pp.

Funky Fungi

Reviewed by **Michael A. Funk**[8]

Whether a strange color sticking out of dull brown leaves, a drooping mane of tendrils attached to a tree, or maybe just an odd funk in the air after a rainy day, fungi are amazing and diverse and a bit strange. In *Funky Fungi*, Alisha Gabriel and Sue Heavenrich convey an infectious love for finding fungi in the wild and explain how these mysterious organisms fit into our ecosystems and lives.

The book is packed with interesting and up-to-date facts about this often-misunderstood kingdom of life. Mushrooms—both those that are edible and those that should not be eaten—are well represented but are just a small sliver of the diversity discussed. Yeasts, molds, insect-zombifying *Cordyceps*, and even a honey mushroom that may be the largest organism on Earth are all on the figurative menu. Speaking of eating, don't expect tips for distinguishing poisonous mushrooms from safe—this book is not a foraging guide and doesn't try to be. Although there is a section on fungi as food, Gabriel and Heavenrich stick to what readers can find in a supermarket. Other ways to appreciate fungi include growing a mold garden, dyeing a piece of fabric with pigments extracted from mushrooms, and starting a compost heap.

In addition to fungus facts, the book includes profiles of mycologists, beautiful high-resolution photos, and lots of activities, both indoor and outdoor. Although some of the terminology and concepts in the text might challenge younger readers, the denser sections are also packed with engaging examples and will hopefully inspire more reading and adventuring. There are a few online resources mentioned near the end and suggestions for field guides if readers want to take their forays to the next level.

Funky Fungi: 30 Activities for Exploring Molds, Mushrooms, Lichens, and More, *Alisha Gabriel and Sue Heavenrich*, Chicago Review Press, 2022, 128 pp.

YOUNG ADULT SCIENCE BOOK

The Complete Guide to Absolutely Everything (Abridged)

Reviewed by **Megan Engel**[11]

With wry, irreverent British humor, Adam Rutherford and Hannah Fry tackle a host of big, albeit seemingly random, questions, devoting a chapter of *The Complete Guide to Absolutely Everything (Abridged)* to each. These range from the age-old and deep (Do we have free will? What would aliens look like?) to the droll (Does my dog love me? Are we all suckers?). Most are not definitively answered, but the quest for solutions is engaging and enjoyable, peppered with pop culture, literary, and classical references and easy-to-digest analogies.

Rutherford and Fry focus less on objective scientific truths and more on us. For example, readers journey through human timekeeping, from sundials to "leap seconds," and learn how the modern financial system hinges precariously on subterranean fiber-optic cables that deliver atomic time to banks.

Instead of Einstein's time dilation, expect to learn about time dilation inside the human body (time seems to slow when we are in peril or cut off from sunlight). And Darwin's theory of evolution takes a backseat to his debunked theories of human emotion.

Despite mostly glossing over technical details, *The Complete Guide* does highlight some fascinating recent scientific findings—the fact that rats appear to feel regret, for example, and a case in which a woman appeared to be able to smell Parkinson's disease on a patient's clothing months before they began exhibiting symptoms—as well as a number of state-of-the-art experiments.

Rutherford and Fry also shine a light on important issues that plague modern science, such as the "File Drawer Problem," wherein "shiny novelties" are prioritized for publication over essential but less flashy experiments that verify existing results, and the related "replication crisis."

Readers are likely to finish *The Complete Guide* with a pocketful of intriguing anecdotes with which to entertain at cocktail parties and a feeling of endearment toward humankind, who dare to ask questions they are "singularly ill-equipped to answer."

For a full-length review of *The Complete Guide to Absolutely Everything (Abridged)*, see "Confronting age-old questions, big and small," *Science* **375**, 35 (2022).

The Complete Guide to Absolutely Everything (Abridged): Adventures in Math and Science, *Adam Rutherford and Hannah Fry*, W. W. Norton, 2022, 304 pp.

Sticky

Reviewed by **Marc S. Lavine**[5]

In *Sticky*, Laurie Winkless takes readers on a tour of what makes some things stick together and others slide past one other, describing how a range of factors can change these interactions. The book includes an exploration of materials such as paints, permanent and temporary adhesives, nonstick coatings, tires, gecko feet, and shark skin. For each, the factors that make them sticky or slippery involve chemical and physical attractions or repulsions. In fact, as Winkless notes, neither "sticky" nor "slippery" are technical terms, and there is no simple numerical scale to delineate between the two states that we often think of as opposites.

Winkless takes a very broad view of surfaces and interfaces, such that the book includes detours to related topics that are often needed to understand the science under

consideration. For example, to understand how air currents interact with the surface of golf balls and airplane wings, the reader requires a crash course in aerodynamics. Likewise, Winkless's discussion of the factors that affect the speed of high-performance race cars goes well beyond the engineering of the car's tires or brakes, describing how specific track and race day conditions affect the chemistry and design of many of the car's components.

We do not often think of slipperiness in the context of earthquakes, observes Winkless, however, it is the stick-slip motion of tectonic plates that leads to cracks in the earth and the movement of faults. Meanwhile, in discussing the surface properties of ice, we learn about a raging scientific debate in the otherwise sedate sport of curling, with two competing theories of the underlying factors that determine the curl of a stone.

Through a wide range of topics, including some that are likely to be less well known, *Sticky* offers readers an insider's guide to the secret science of surfaces.

Sticky: The Secret Science of Surfaces, *Laurie Winkless*, Bloomsbury Sigma, 2022, 336 pp.

Astroquizzical

Reviewed by **Adrian Cho**[12]

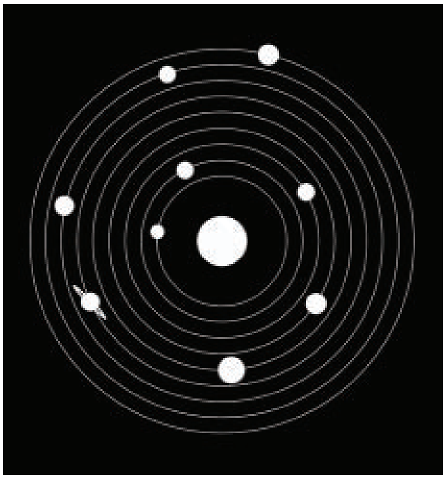
Ever wonder why stars twinkle, what will happen to Earth when the Sun dies, or how we know that the Universe is expanding? This lavishly illustrated, fact-packed book answers these questions and more in a way that is both accessible and conceptually meaty. The reader comes away with a solid introduction to astronomy, planetary science, stellar and galactic astrophysics, and cosmology.

The reader takes a tour of the cosmos guided by Jillian Scudder, an astrophysicist at Oberlin College and Conservatory, who since 2013 has fielded questions on her blog of the same title. Moving ever outward, she opens with the night sky, takes a close look at the Moon, and swings through the planets in our solar system. She examines the Sun and other stars, zooms out to the galaxies, and finally contemplates the Universe as a whole.

Among the questions, Scudder sprinkles 10 thought experiments, musings on some hypothetical scenario, such as, "What would happen if we split the Sun in half?" Some of these digressions are pretty quirky. Scudder asks, for example, what would happen if you could open a door on Earth's surface and step out on the Moon? The answer: nothing good, thanks to the 1480 kilometer per hour wind that would spit you through the door.

The real stars of the show are the cosmos's own treasures. The book features dozens of jaw-dropping photos of marvels such as Olympus Mons, the 24-kilometer-tall volcano on Mars, and the fiery face of the Sun speckled with coronal mass ejections. The famous photo of countless galaxies taken by the Hubble Space Telescope may well bring the viewer to an awestruck stop. Thanks to Scudder's accessible prose and the luscious illustrations, readers will come away with many answers, but likely even more questions to fuel their curiosity.

Astroquizzical: Solving the Cosmic Puzzles of our Planets, Stars, and Galaxies (The Illustrated Edition), *Jillian Scudder*, Icon Books, 2021, 224 pp.



The Last Days of the Dinosaurs

Reviewed by **Victoria Arbour**[13]

Sixty-six million years ago, the beginning of the end of the age of dinosaurs came crashing down in the form of an asteroid that struck what is now the Yucatán Peninsula in Mexico. The details of this cataclysmic event, and what happened afterward, have rarely received the feature-length treatment that Riley Black presents in *The Last Days of the Dinosaurs*.

In just over 300 pages, Black pulls together decades of scientific research on the demise of the nonavian dinosaurs into a deeply compelling narrative of both luck and misfortune in the face of almost unimaginable calamity. Starting just a few days before the asteroid impact, Black centers the story on the animals present in the Hell Creek Formation of the western United States, a geological unit that has been studied for more than 100 years and that provides the best window we have into the time before and after the mass extinction. Through the eyes of the inhabitants of Hell Creek, readers pass through the moment of impact and then the first hour, day, month, and year of the Cenozoic era.

The story continues with chapters set one hundred, one thousand, one hundred thousand, and one million years after the asteroid impact, charting the reshaping of the world. Each chapter ends with a detour to somewhere else on the planet—Antarctica, India, and the Atlantic Ocean, for example—providing a global balance to this otherwise tightly focused narrative.

Unlike the other mass extinctions recorded in the fossil record, extinction for most species at the end of the Cretaceous probably happened within a few hours or days after the asteroid impact. Black's writing captures the horror of the impact event without lingering on doom and gloom, instead keeping a steady momentum focusing on survival and change.

Black recounts stories of lucky survivors whose descendants will be familiar to many readers—frogs, snakes, turtles, and crocodiles—and organisms that rarely get a mention in dinosaur stories, such as the planktonic coccoliths and coil-shelled ammonites. And of course, as the nonavian dinosaurs pass into history, we see the changes that happen to the ancestors of today's mammals, who evolved into new forms in the asteroid's aftermath. ■

For a full-length review of *The Last Days of the Dinosaurs*, see "Inside the dinosaurs' demise," *Science* **376**, 360 (2022).

The Last Days of the Dinosaurs: An Asteroid, Extinction, and the Beginning of Our World, *Riley Black*, St. Martin's Press, 2022, 304 pp.

10.1126/science.adf7725

ILLUSTRATION: ROBERT BRANDT

[1]The reviewer is a deputy editor at *Science*. Email: svignier@aaas.org [2]The reviewer is the Letters editor at *Science*. Email: jsills@aaas.org [3]The reviewer is a senior correspondent at *Science*. Email: jmervis@aaas.org [4]The reviewer is a senior editor at *Science*. Email: cash@science-int.co.uk [5]The reviewer is a senior editor at *Science*. Email: mlavine@aaas.org [6]The reviewer is a senior editor at *Science*. Email: sscanlon@science-int.co.uk [7]The reviewer is a senior editor at *Science*. Email: ynusinov@aaas.org [8]The reviewer is a senior editor at *Science*. Email: mfunk@aaas.org [9]The reviewer is a senior editor at *Science*. Email: phines@aaas.org [10]The reviewer is a senior editor at *Science*. Email: vthompso@aaas.org [11]The reviewer is at the School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA. Email: mcengel@seas.harvard.edu [12]The reviewer is a news writer at *Science*. Email: acho@aaas.org [13]The reviewer is curator of paleontology at the Royal BC Museum, Victoria, BC V8W 9W2, Canada. Email: victoria.arbour@gmail.com



Psychedelic-assisted psychotherapy is one model of psychedelic medicine that could be used to treat some neuropsychiatric disorders.

NEUROSCIENCE

The therapeutic potential of psychedelics

The development of psychedelics as medicines faces several challenges

By **Emmanuelle A. D. Schindler**[1,2] *and* **Deepak Cyril D'Souza**[3,4]

Psychedelics are reported to have rapid-onset and long-lasting therapeutic benefits after a single or few doses. Sustained (1 year or more) clinical benefits have been reported in depression and smoking cessation studies after just two or three doses of psilocybin combined with psychotherapy (*1*, *2*). By contrast, conventional medications for neuropsychiatric disorders take days to weeks to begin working and need to be taken daily over prolonged periods and sometimes indefinitely. Other potential applications for psychedelics include treatment of cancer-related anxiety, obsessive-compulsive disorder, headache disorders, and phantom limb syndrome. Although arguably paradigm shifting, a number of unanswered questions remain about psychedelics as medicines, including the definition of a psychedelic drug, the mechanism of therapeutic effects, optimizing clinical benefit, and verifying safety.

[1]Neurology Department, Yale School of Medicine, New Haven, CT, USA. [2]Headache Center of Excellence, VA Connecticut Healthcare System, West Haven, CT, USA. [3]Psychiatry Department, Yale School of Medicine, New Haven, CT, USA. [4]Neuropsychiatry Firm, VA Connecticut Healthcare System, West Haven, CT, USA. Email: emmanuelle.schindler@yale.edu; deepak.dsouza@yale.edu

Lysergic acid diethylamide (LSD), psilocybin, *N*,*N*-dimethyltryptamine (DMT), 2,5-dimethoxy-4-iodoamphetamine (DOI), and mescaline are some of the agents collectively categorized as classic psychedelics. These drugs are all agonists at 5-hydroxytryptamine (5-HT; serotonin) 2A (5-HT_{2A}) receptors and produce characteristic acute psychedelic effects, which include alterations in perception, feeling, and consciousness (*3*). Although other compounds produce psychedelic-like effects as well as clinical benefits [such as ketamine and 3,4-methylenedioxymethamphetamine (MDMA; "ecstasy")], they are pharmacologically distinct, and so the focus here is on classic psychedelics.

Narrowly defining a drug class by one set of its effects (psychedelic) can be problematic because it colors the perception and may ultimately limit the breadth of its application. For example, patients may be hesitant to take an antidepressant for a nonpsychiatric condition, such as peripheral neuropathy or migraine, simply because of the class name. Alternate terms offered for psychedelics include "psychoplastogens" or "neuroplastogens" (*4*), which remove prejudice and highlight the ability of these drugs to induce change, although not necessarily the distinct dosing regimen. Borrowing from headache medicine, transitional medications are those taken for a short time and that suppress headache for a prolonged period well beyond the treatment itself (steroid pulse). A compound term such as "transitional neuroplastogen" captures the notions of long-lasting change after a brief treatment period.

The mechanism of therapeutic effects of psychedelics is widely queried but remains unclear. 5-HT_{2A} receptor antagonists block acute psychedelic effects, but whether they also block therapeutic effects requires further investigation. To what extent the many other direct or indirect targets of psychedelics—such as 5-HT_{1A}, 5-HT_{2B}, and 5-HT_{2C} receptors; dopamine receptors; α-adrenergic receptors; monoamine transmission; and glutamatergic transmission—contribute to therapeutic effects is also not known (*3*). Once bound to a receptor, a ligand may also activate one or more intracellular processes. For example, the β-arrestin signaling pathway has been suggested to be relevant for antidepressant effects of 5-HT_{2A} receptor activation but not psychedelic effects (*5*). Psychedelics also have numerous physiological effects, including anti-inflammatory, hormonal, and epigenetic effects, which have pathological relevance in such conditions as depression, substance abuse, and headache disorders (*6*).

How any of these transient effects on receptors or biological systems might explain sustained therapeutic effects is unknown. The initiation of a cascade of events with en-

PHOTO: BAZA PRODUCTION/SHUTTERSTOCK

during neuroplastic effects at the cellular and network level is one plausible and popular theory. Classic psychedelics have been shown in cellular and in vivo preclinical models to promote synaptogenesis and increase cortical dendritic spine size, number, and complexity (*7*), with some effects lasting a month (*8*). In pigs, a single intravenous dose of psilocybin was shown to induce lasting (7 days) increases in cortical and hippocampal synaptic density (*9*). Interestingly, experiments identified some, but not other, cellular changes to be 5-HT_{2A} receptor-mediated. Whether these neuroplastic cellular changes are related to durable therapeutic effects could be investigated in models of human disease or human patients. For example, in patients with treatment-resistant depression, changes in brain resting-state functional connectivity the day after completing a two-dose regimen of psilocybin correlated with a lasting clinical improvement at 5 weeks (*10*). Moreover, various changes in brain connectivity persisted for 1 month after a single dose of psilocybin, as did increases in positive mood and decreases in anxiety (*11*). These postpsychedelic connectivity changes suggest an association with and perhaps a source for therapeutic effects. However, replication of findings in placebo-controlled studies, over a longer term, and with clearly delineated modeling and analytical parameters is necessary to show this more conclusively. In addition, comparisons with other drugs that also induce neuroplastic changes, such as ketamine (*8*), are necessary to characterize signature effects of psychedelics. Studies in different patient populations will also be required to identify the changes relevant to specific disorders, such as hypothalamic function in cluster headache (*6*).

The neuroplastic effects of psychedelics may serve to open a therapeutic window, allowing other drugs or treatments to take effect. For example, psychedelic drug studies in depression include a course of psychotherapy, a standard treatment used in depression. In a case of phantom limb syndrome, psilocybin mushrooms were reported to have synergistic and lasting therapeutic effects when used in conjunction with mirror therapy, a standard rehabilitative therapy used to reverse aberrant somatosensory cortex reorganization in that condition (*12*). These psychedelic-assisted therapies use the drug with an existing disease-specific therapy. However, psychedelics may also have independent treatment effects. Cluster headache patients have been self-administering them as standalone treatment for decades, and clinical trials in headache disorders have modeled this method. Ultimately, the independent and interactive effects of psychedelics with other disease-specific therapies must be investigated systematically. The type and duration of adjunctive treatment must also be considered (which form of psychotherapy for depression). The need to repeat drug treatment at certain intervals is also anticipated, although existing clinical trials are relatively short. Furthermore, some study protocols include curated decoration, music, and so on during drug dosing, as well as the presence of therapists to guide or enhance the experience. The specific settings and interactions that are necessary and optimal for therapeutic outcomes have not been systematically examined. Notably, whether additional treatments and procedures can be implemented on a large scale and reach all populations in need must be factored into the development of psychedelics as medicines.

Positive correlations between the magnitude of the psychedelic experience and therapeutic benefit have not been consistently observed. In headache disorders, acute psychedelic effects appear unrelated to therapeutic outcomes (*13*). Across studies, the scales (and subscales) used to measure psychedelic effects are not aligned, leaving the relationship between acute subjective effects and clinical effects unclear. Furthermore, different psychedelics produce distinct acute experiences. Seeking to understand the origin of specific acute perceptual and other subjective effects and their relevance in treating particular neuropsychiatric conditions could further optimize treatment. This could be done through a number of complementary experimental manipulations: comparing the therapeutic efficacy of subpsychedelic versus psychedelic doses, conducting wide dose-response studies, blocking the psychedelic effects with targeted receptor antagonists, using analogs that lack psychedelic effects [such as 2-bromo-LSD (BOL)], comparing classic psychedelics or using other psychotropic drugs with distinct pharmacological profiles (such as MDMA), or administering drugs to individuals while they are in natural or induced sleep. If the acute psychedelic effects of these drugs are central to some therapeutic effects, it will be critical to determine what level and duration are necessary. For example, intravenous DMT showed next-day (rapid) antidepressant effects and produced acute psychedelic effects for ~30 minutes (*14*), contrasting oral psilocybin's 6 hours or LSD's 12 hours of acute effects. A shorter psychedelic experience, if resulting in the same clinical benefit, would be more logistically feasible and palatable.

"The neuroplastic effects of psychedelics may serve to open a therapeutic window..."

Unlike the development of drugs in the pharmaceutical industry, there is a massive amount of information about psychedelics available to the public before their implementation as medicines. Media coverage is not bound by the standards of accuracy of scientific reporting and seems biased toward covering the "universally life-changing" abilities of psychedelics. This raises expectations for success in clinical trials, which may only partially be tempered by education (*15*). Study results are also affected by potential unblinding from the unmistakable acute effects of psychedelics and lack of acute effects with placebo (*15*). An active control that produces acute subjective effects may minimize unblinding, although identifying such an agent is challenging. A low dose of the psychedelic being investigated may be used, but this too may produce lasting therapeutic effects (*13*). A related drug with overlap in several dimensions of the psychedelic experience, such as ketamine, may be tried, although ketamine also has lasting clinical effects. The use of other nonclassic psychedelics with no known therapeutic effects (such as salvinorin A) might be considered. Other control agents that have been used include niacin and diphenhydramine, although these drugs do not entirely substitute for acute effects, particularly of higher doses of psychedelics. Additional methods for maintaining blinded conditions include recruiting psychedelic-naïve subjects, emphasizing interpersonal variability in the acute effects, and incomplete disclosure about drugs or doses that may be received (*15*).

Given the increase in psychedelic research, together with the surge in popular and commercial interests, the safety of psychedelics must not only be revisited but considered in the context of current and future use. Historically, the use of psychedelics has involved the infrequent consumption of moderate to high doses. In research, limited dosing (single or a few doses) is studied, and the drug is administered under controlled conditions with medical and psychiatric oversight to carefully screen and prepare participants. Such practices support safety and tolerability and deter misuse. However, practices that deviate from this model are emerging. One version of the practice of "microdosing" involves repeated exposure to low or subperceptual doses of a psychedelic over a prolonged period. Although "micro" might sound appealing and denote safety, there is no evidence that the frequent and long-term use of psychedelics (at any dose) is safe. As a case in point, the LSD-derivative methysergide, an effective migraine and cluster headache preventive (taken daily), was removed from the market after cases of cardiac valve fibrosis and other tissue fibrosis emerged.

The fibrogenic effects are related to 5-HT_{2B} receptor activation, and although psychedelics have varying affinities for this receptor (highest for ergot derivatives), frequency and duration of exposure must be considered in the pharmacodynamics of these new, unverified regimens.

Psychedelics may also have acute therapeutic effects (for example, aborting a headache attack). Although potentially acceptable for conditions that require infrequent use, the frequent consumption of these drugs for the acute management of a chronic and/or persisting condition is not only impractical but risks tolerance and loss of efficacy and has not been systematically studied for safety. Indeed, an ongoing challenge within pain management is the reliance on abortive rather than preventive treatment, which leads to sensitization and dependency. Psychedelics have historically failed to demonstrate addictive properties, but the neuropsychological impact of frequent (and potentially increasing) use needs further study. In addition to pharmacology and purpose of use, other factors that contribute to how a drug is used (or misused) include availability, perception, commercialization, and promotion. Furthermore, the idea that psychedelics may be used outside of a diagnosed medical condition—say, for general life enhancement or improved concentration—is intriguing but will also require formal investigation. Without dedicated study, new regimens and applications may have unexpected outcomes. The comprehensive investigation of psychedelics and their implementation as legitimate medicines remain valuable but substantial undertakings. ■

REFERENCES AND NOTES

1. N. Gukasyan *et al.*, *J. Psychopharmacol.* **36**, 151 (2022).
2. M. W. Johnson, A. Garcia-Romeu, R. R. Griffiths, *Am. J. Drug Alcohol Abuse* **43**, 55 (2017).
3. D. E. Nichols, *Pharmacol. Rev.* **68**, 264 (2016).
4. R. Strassman, *The Psychedelic Handbook: A Practical Guide to Psilocybin, LSD, Ketamine, MDMA, and Ayahuasca* (Ulysses Press, 2022).
5. D. Cao *et al.*, *Science* **375**, 403 (2022).
6. E. A. D. Schindler, R. M. Wallace, J. A. Sloshower, D. C. D'Souza, *Front. Pharmacol.* **9**, 177 (2018).
7. C. Ly *et al.*, *Cell Rep.* **23**, 3170 (2018).
8. L. X. Shao *et al.*, *Neuron* **109**, 2535 (2021).
9. N. R. Raval *et al.*, *Int. J. Mol. Sci.* **22**, 835 (2021).
10. R. L. Carhart-Harris *et al.*, *Sci. Rep.* **7**, 13187 (2017).
11. F. S. Barrett, M. K. Doss, N. D. Sepeda, J. J. Pekar, R. R. Griffiths, *Sci. Rep.* **10**, 2214 (2020).
12. V. Ramachandran, C. Chunharas, Z. Marcus, T. Furnish, A. Lin, *Neurocase* **24**, 105 (2018).
13. E. A. D. Schindler *et al.*, *Neurotherapeutics* **18**, 534 (2021).
14. D. C. D'Souza *et al.*, *Neuropsychopharmacology* **47**, 1854 (2022).
15. J. S. Aday *et al.*, *Psychopharmacology* **239**, 1989 (2022).

ACKNOWLEDGMENTS

The authors receive research funding from Heffter Research Institute, Ceruvia Lifesciences, and Wallace Research Foundation. They serve on the scientific advisory boards of Ceruvia Lifesciences and/or Clusterbusters and are named as inventors on patent US20210236523A1.

10.1126/science.abn5486

PHOTO: JJ GOUIN/ALAMY STOCK PHOTO



Research using herbarium specimens collected over nearly 200 years demonstrates how the native North American weed waterhemp (shown here invading a soybean field) adapted to agricultural practices over space and time.

AGRICULTURE

Herbaria reveal cost of the Green Revolution

Rapid weed evolution is exposed by genome sequencing of natural history collections

By **Katherine Waselkov**[1] *and* **Kenneth M. Olsen**[2]

A visitor to the US Midwest will be immediately struck by the sheer scale of industrially farmed corn and soybean. These fields are intensively managed artificial ecosystems, from their planting and harvesting timelines to the fertilizers and pesticides that are continually applied. Evolutionary biologists have long presumed that weeds are under strong natural selection to adapt to this anthropogenic ecosystem, which first appeared in the mid–20th century's "Green Revolution" in agriculture. On page 1079 of this issue, Kreiner *et al.* (*1*) report that the selection pressure on weeds in modern agricultural fields is higher than estimates from most other natural systems (*2*). The authors leverage historical samples in natural history collections to temporally link the adaptation of the agricultural weed waterhemp to the Green Revolution.

The intensification of agriculture in the US and Canada has resulted in higher crop yields on less acreage in exchange for greater use of nitrogen-based fertilizers and pesticides. However, the maintenance of crop monocultures in these biodiversity-poor ecosystems spurred an arms race against weeds, insect pests, and microbial pathogens. To date, 267 plant species have evolved resistance to at least one chemical group of herbicides (a type of pesticide) meant to control their presence in agricultural fields (*3*). This intense, human-mediated selection pressure increased in the late 1990s with the widespread adoption of soy, cotton, and corn varieties that were genetically modified to resist glyphosate pesticides. This chemical (commercially known as Roundup) is currently the most popular herbicide in the US (*4*). Early fears about the escape of modified genes from genetically modified crops have largely been allayed through careful design and regulation. However, the proliferation of weeds that have evolved resistance to glyphosate largely through naturally occurring mutations has been an unanticipated consequence of the commercialization of glyphosate-resistant crops (*5*). As the continuous, exclusive use of Roundup leads to the emergence of more glyphosate-resistant weed populations and species (currently 56), farmers are resorting to older chemicals or more expensive weed-control methods. One of the most pervasive and damaging glyphosate-resistant weed species is the North American native waterhemp (*Amaranthus tuberculatus*).

Waterhemp is unusual among agricultural weeds in that individual plants are either male or female and thus must cross-pollinate to reproduce (unlike many weeds that self-pollinate). Encountering a mate is enabled by wind pollination and enormous popula-

[1]Department of Biology, California State University, Fresno, Fresno, CA 93740, USA. [2]Department of Biology, Washington University, St. Louis, MO 63130, USA. Email: kwaselkov@csufresno.edu; kolsen@wustl.edu

tion sizes as well as a staggering reproductive capacity in optimal conditions (up to 1 million seeds per female plant) (*6*). Waterhemp's native range is from the US Great Plains to southern Ontario, Canada, where its original habitat was the periodically flooded margins of bodies of water. Reconstructing the historical invasion of wild waterhemp into agricultural fields is complicated by the existence of two taxonomic varieties within the species (*7*). The variety *rudis* in the western part of the species range (west of the Mississippi River) had preexisting characteristics that allowed it to colonize human-disturbed environments. Its human-assisted migration eastward into the range of the variety *tuberculatus* began in the mid-1800s, marking the origin of the agricultural weed (*8*). Waterhemp was not a major agricultural problem until the 1990s, when it began to evolve resistance to herbicides. To date, various US populations of waterhemp have evolved resistance to seven different chemical classes of herbicides, and many populations show resistance to multiple chemicals (*3*).

The study of Kreiner *et al.* is the culmination of several previous studies on waterhemp evolution (*9*–*11*). The authors used whole-genome sequence data to identify all genes with substantial differences between geographically paired agricultural and natural populations. The resulting candidate genes for agricultural adaptation will provide a valuable starting point for future studies aimed at identifying and functionally confirming genes and mutations involved in weedy traits (including herbicide resistance and other as-yet-unknown traits for agricultural adaptation). Unexpectedly, their estimates of very strong selection on three common herbicide-resistance genes are in the opposite direction predicted by most previous literature. The fate of a resistance mutation in a population is determined in part by its fitness. Mutations that suffer little or no fitness cost are more likely to persist in the absence of herbicide. The fitness costs of being herbicide resistant in natural environments appear to be greater than the benefits of being herbicide resistant in agricultural environments. The fitness costs of carrying an herbicide-resistance mutation in the absence of herbicides depend on the particular mutation, the other genes present in the genome, and the other selection pressures in the environment. The estimates of Kreiner *et al.* are the first to be based on empirical measurements in a nonexperimental context (*12*).

The most interesting aspect of the study of Kreiner *et al.* is the detection of temporal changes in agriculture-associated gene frequencies and the measurement of selection over time. These inferences were made possible by DNA extraction and whole-genome sequencing from herbarium specimen samples, including those dating back to the mid–19th century. Many evolutionary studies have used DNA from museum specimens, including research on weeds, invasive species, and herbicide resistance (*13*, *14*), but the scale and quality of the genetic data analyzed by Kreiner *et al.* have set the standard for future research. The authors followed best practices by extracting DNA in a laboratory designed to handle ancient DNA. Although herbarium-derived data about plant populations must be interpreted with caution—given the necessarily limited sample size per time period and the known biases in plant collection (*13*)—short-read whole-genome sequencing technology has made it possible to analyze fragmented DNA, such as the degraded genetic material from older specimens. The judicious use of destructive sampling of museum specimens will play an ever-larger role in the study of evolutionary biology. Unfortunately, museums are underfunded and collection efforts are decreasing at a time when having whole specimens for future unknown uses is more important than ever given the rate of biodiversity loss (*13*).

Kreiner *et al.* show that waterhemp adapted to agricultural intensification through migration and selection of preexisting genetic variation, followed by the rapid evolution of herbicide resistance through new mutations. As industrial-scale agriculture becomes more mechanized and technology driven (*15*), new selection pressures will spur the rise of new adaptations and weed species in crop fields. Future evolutionary biologists will rely on continued collection of herbarium specimens, most likely analyzed with currently unavailable technology, to understand the ongoing arms race between weeds and agricultural pest management. ■

REFERENCES AND NOTES

1. J. M. Kreiner *et al.*, *Science* **378**, 1079 (2022).
2. T. J. Thurman, R. D. H. Barrett, *Mol. Ecol*. **25**, 1429 (2016).
3. I. Heap, The International Herbicide-Resistant Weed Database, www.weedscience.org [accessed 1 September 2022].
4. S. O. Duke, S. B. Powles, *AgBioForum* **12**, 346 (2009).
5. National Academies of Sciences, Engineering, and Medicine, *Genetically Engineered Crops: Experiences and Prospects* (The National Academies Press, 2016).
6. L. E. Steckel, *Weed Technol*. **21**, 567 (2007).
7. S. L. Mosyakin, K. R. Robertson in *Flora of North America North of Mexico*, Flora of North America Editorial Committee, Ed. (Oxford Univ. Press, 2003), pp. 410–435.
8. J. Sauer, *Evolution* **11**, 11 (1957).
9. J. M. Kreiner *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 21076 (2019).
10. J. M. Kreiner *et al.*, *eLife* **11**, e70242 (2022).
11. J. M. Kreiner *et al.*, *Evolution* **76**, 70 (2022).
12. M. M. Vila-Aiub, *Plants* **8**, 469 (2019).
13. P. L. M. Lang *et al.*, *New Phytol*. **221**, 110 (2019).
14. C. Délye *et al.*, *PLOS ONE* **8**, e75117 (2013).
15. C.-X. Zeng *et al.*, *Plant Methods* **14**, 43 (2018).

ACKNOWLEDGMENTS

We thank the Olsen Lab at Washington University for input.

10.1126/science.ade4615

DRUG DEVELOPMENT

Getting bifunctional molecules into cells

A class of transmembrane proteins helps shuttle large drugs across the cell membrane

By **R. Scott Lokey**[1] *and* **Cameron Pye**[2]

Simultaneously optimizing biochemical potency and membrane permeability remains one of the biggest challenges in drug development. For typical drug targets, such as enzymes and receptors, finding a configuration of atoms that will allow target binding with high affinity can be relatively easy to achieve using drug-like molecules, which are generally below ~600 Da. However, there is a long list of therapeutically compelling targets inside the cell that either lack the kinds of deep pockets appropriate for binding drug-like small molecules or have multiple pockets separated over large distances. For molecules that are large enough to effectively engage challenging protein interfaces, getting past the cell membrane can be insurmountable. On page 1097 of this issue, Lou *et al.* (*1*) describe a mechanism by which large, dumbbell-shaped (bitopic) molecules with molecular weights as high as 1784 Da can access the cytoplasm, revealing a potential avenue to reach intractable intracellular targets.

Most small molecules cross membranes by diffusion (*2*). This passive mechanism of membrane transport requires striking a balance between membrane permeability and other competing factors. If a molecule is too hydrophilic, it cannot partition into the membrane, whereas if it is too hydrophobic, it will tend to aggregate in aqueous compartments or get stuck in the membrane. Although navigating these constraints with drug-like small molecules is commonplace, the rules that govern passive permeability become less clear when molecules become larger (*3*, *4*).

[1]Department of Chemistry and Biochemistry, University of California Santa Cruz, Santa Cruz, CA, USA. [2]Unnatural Products, Santa Cruz, CA, USA. Email: slokey@ucsc.edu; cameron.pye@unnaturalproducts.com

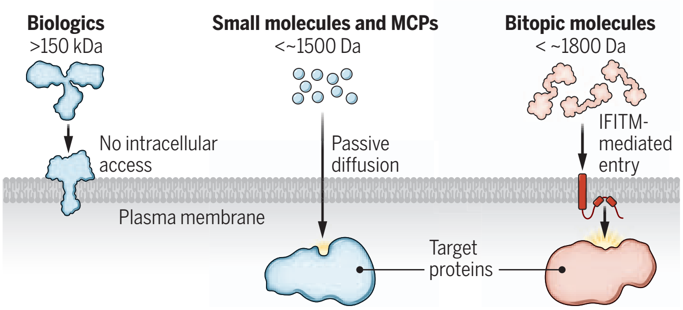
Large macrocyclic peptides (MCPs) from natural products serve as counterexamples to the long-held assumption that the passive mechanism is only available to low–molecular weight drugs. Cyclosporine A (CSA), an immunosuppressive drug with a molecular weight of >1200 Da, maintains passive cell permeability and oral bioavailability despite its size (*5*). Early examples of fully synthetic MCPs are making their way into the clinic (*6*), and more will follow as the tools and rules to enable the design of MCPs with passive permeability are refined. However, for a growing number of large, bitopic molecules, like those described by Lou *et al.*, their cellular potencies are difficult to reconcile with often very low passive permeabilities. Such cases suggest the possibility of more active modes of transport through cell membranes, raising the question as to which proteins, if any, are involved.

Lou *et al.* take a functional genomics approach. They used a genome-wide CRISPR screen in patient-derived chronic myeloid leukemia cells to identify proteins whose expression is required for the cellular effect of RapaLink-1, a bitopic inhibitor of mammalian target of rapamycin (mTOR) composed of the allosteric binder rapamycin covalently attached to the kinase active site inhibitor sapanisertib by way of a long, flexible linker. Despite its high molecular weight of nearly 1800 Da, RapaLink-1 shows unexpected cellular efficacy and is even able to traverse the blood-brain barrier, which is more limiting than the gut wall or the membranes of isolated cells. Their screen revealed a strong chemical-genetic relationship between the cellular efficacy of RapaLink-1 and the expression of interferon-induced transmembrane proteins (IFITMs). Furthermore, the sensitivities of hundreds of cell lines to RapaLink-1 were highly correlated to IFITM expression. Combined with the hypothesized role of IFITMs in blocking viral entry by modulating membrane properties at the virus–host cell interface (*7*), these observations suggest a role for IFITM proteins in mediating bitopic drug cell entry (see the figure).

The authors also investigated the effect of IFITM expression on the potency of other bitopic molecules unrelated to RapaLink-1. Lou *et al.* generated an inhibitor of the oncogenic kinase breakpoint cluster region (BCR)–ABL1 called DasatiLink-1, in which they combined dasatinib, a well-established clinical inhibitor of the kinase active site, with asciminib, an allosteric inhibitor that binds to the myristoyl pocket, which is distant from the active site. They also produced a covalently cross-linked dimer of the natural product rocaglamide, which binds the eukaryotic initiation factor 4A-I (EIF4A1) helicase and inhibits translation. Not only were DasatiLink-1 and BisRoc-1 active in cells, but, as with RapaLink-1, the potency of both drugs was significantly enhanced by IFITM expression. Although the cellular potency of the rocaglamide dimer (BisRoc-1) was similar to that of the unlinked molecule—which was also the case for the individual component drugs that comprise RapaLink-1 and DasatiLink-1—reducing IFITM expression diminished the efficacy of the dimer but not rocaglamide alone.

Gaining entry to cells

Membrane permeability is a major challenge in drug development. Biologics, such as antibodies, can only target extracellular protein domains and soluble targets. Macrocyclic peptides (MCPs) from natural products can passively diffuse across the cell membrane. Small molecules can be designed to passively traverse the cell membrane, but their size limits the region they can bind in target proteins. Bitopic molecules can engage much larger regions of intracellular proteins, broadening potential targets, and their entry into cells is at least partially mediated by interferon-induced transmembrane proteins (IFITMs).



The work by Lou *et al.* provides a glimpse into a potential mechanism by which large, bitopic molecules can enter cells, but the true importance of IFITM proteins in mediating the cellular effects of these molecules awaits further study. For all three of the bioactive bitopic molecules described, the differences in potency when the three *IFITM* family members were deleted compared with wild-type cells were modest: 6-fold for RapaLink-1, 2.5-fold for DasatiLink-1, and 2.2-fold for BisRoc-1. This raises questions about the residual activity of these molecules in IFITM-deleted cells: Are there other, protein-mediated mechanisms by which these molecules get into cells, or is the effect mediated by IFITMs indirect? Perhaps another CRISPR screen using *IFITM*-deleted cells could reveal additional players. It is also possible that these molecules exhibit a limited degree of passive permeability, as has been shown with other highly potent bitopic molecules (*8*). The extent to which IFITM proteins are required for the membrane transport of such molecules, and the structural requirements for IFITM-mediated transport, will be exciting areas for further study.

Although it is inspiring to see an additional mechanism of cell entry start to be elucidated, relying on IFITM-mediated transport in the therapeutic context could introduce complexities and challenges. The expression of IFITM proteins is likely to vary between patients and even among cell types within the same patient. Furthermore, for anticancer compounds that require IFITMs for therapeutic efficacy, down-regulation of IFITMs could emerge as a mechanism of acquired resistance.

There is clearly a rich landscape of bioactivities in large, multifunctional molecules. Natural products have shown the promise of breaking the rules of traditional small-molecule medicinal chemistry, but the challenges associated with achieving their biochemical complexities while balancing drug-like properties have long placed limits on efforts to mimic their success with purely synthetic molecules. The excitement around targeted protein degraders [proteolysis-targeting chimeras (PROTACs)] and other induced proximity strategies over the past decades has at least partially been driven by the modular, designable nature of these molecules, whereby two distinct and unrelated binders can be linked together to achieve an entirely new biological effect. Criticisms of these molecules often focus on their atypical molecular structures and the accompanying poor biophysical properties, but Lou *et al.* pave the way for further understanding the complexities of cell entry and bioactivity of this rapidly evolving class of therapeutic molecules. ■

REFERENCES AND NOTES

1. K. Lou *et al.*, *Science* **378**, 1097 (2022).
2. A. Missner, P. Pohl, *ChemPhysChem* **10**, 1405 (2009).
3. S. M. Linker *et al.*, *RSC Adv.* **12**, 5782 (2022).
4. A. A. Rzepiela, L. A. Viarengo-Baker, V. Tatarskii, R. Kombarov, A. Whitty, *J. Med. Chem.* **65**, 10300 (2022).
5. C. K. Wang, J. E. Swedberg, P. J. Harvey, W. Kaas, D. J. Craik, *J. Phys. Chem. B* **122**, 2261 (2018).
6. K. Nomura *et al.*, *J. Med. Chem.* **65**, 13401 (2022).
7. J. M. Perreira, C. R. Chin, E. M. Feeley, A. L. Brass, *J. Mol. Biol.* **425**, 4937 (2013).
8. Y. Atilaw *et al.*, *ACS Med. Chem. Lett.* **12**, 107 (2021).

ACKNOWLEDGMENTS

C.P. is the CEO of Unnatural Products, Inc., which focuses on macrocycle drug discovery. R.S.L. is a consultant for Unnatural Products, Inc.

10.1126/science.adf4412

GRAPHIC: A. FISHER/*SCIENCE*

COMPUTER SCIENCE

AlphaCode and "data-driven" programming

Is ignoring everything that is known about code the best way to write programs?

By **J. Zico Kolter**[1,2]

Competitive programming problems represent a challenging task for even skilled programmers: Given a short natural language description of an algorithmic problem, contestants must quickly write a program that solves the task. On page 1092 of this issue, Li *et al.* (*1*) present the AlphaCode system, which represents a substantial step forward in the development of machine learning (ML) models that can synthesize computer programs to solve these types of challenging problems. But what is perhaps most surprising about the system is what AlphaCode does not do: AlphaCode contains no explicit built-in knowledge about the structure of computer code. Instead, AlphaCode relies on a purely "data-driven" approach to writing code, learning the structure of computer programs by simply observing lots of existing code.

AlphaCode was built by using a type of ML model called a "large language model" (LLM), which operates by predicting text one "token" at a time (tokens are collections of a small sequence of characters, but they can be thought of as single characters in the code). AlphaCode predicts the first character in the solution code (given the problem description), then the second character (given the description and first character), the third character (given the description and first and second characters), and so on. Recent work has highlighted the potential for LLMs to generate code—for example, evaluations of the GPT-3 system (*2*) for code creation and the release of GitHub's Copilot using the Codex system (*3*)—but AlphaCode's ability to generate entire programs that solve competitive programming problems is a notable advance. Although AlphaCode contains some architectural differences from this past work (for example, AlphaCode does not need to generate problem descriptions, so it uses a slightly different mechanism to encode the problem text), the underlying system operates in a manner similar to previous LLMs.

What fundamentally makes AlphaCode outperform other systems on the competitive programming task boils down to two main attributes: training data and postprocessing of candidate solutions. Arguably, the most important lesson for any ML system is that it should be trained on data that are similar to the data it will see at runtime. For this purpose, the AlphaCode developers built the CodeContests dataset, building on two existing datasets (*4*, *5*). CodeContests consists of the explanatory text of ~13,500 competitive programming problems, some simple test cases of desired input-output pairs, and many potential solutions for each problem across several programming languages. The system is also trained on unstructured code from GitHub (characters of code in many languages, without a problem description).

However, generating a single piece of solution code for a problem performs poorly; writing code token by token is prone to mistakes by writing invalid code or programs that produce incorrect results. Thus, AlphaCode also generates thousands of candidate solutions and filters them by ensuring that they are valid and pass simple test cases. The system also clusters similar solutions and submits a single example from each cluster to increase the diversity of the potential solutions.

"It may seem surprising that this procedure has any chance of creating correct code."

What is notably missing from the AlphaCode system is any architectural design in the ML model that relates to the task of interest: producing code. Computer code is a highly structured medium; programs must adhere to a defined syntax and must produce well-defined pre- and postconditions within the different portions of the solution. Program synthesis has a long history (*6*) and numerous techniques have been developed for generating programs that obey these types of constraints. It seems only natural that given a medium as structured as computer code, this structure would be used by ML models aiming to write code.

But AlphaCode does none of this. It generates code the way LLMs generate any text—one token at a time—and only checks for program correctness after the entire program has been written. It may seem surprising that this procedure has any chance of creating correct code. But the reality is that given the proper data and model complexity, coherent structure can emerge. The ultimate recipe for this sequential generation procedure is buried inscrutably within the parameters of the LLM. There are notable limitations to this general approach: Models can sometimes effectively memorize small patterns within the dataset and then output such patterns in rearranged fashion, leading some to term them "stochastic parrots" (*7*)—systems with no real understanding of the underlying question but that are able to blindly mimic likely outputs. Yet AlphaCode is not merely copying existing solutions: Li *et al.* examine and rule out this possibility. Ultimately, AlphaCode performs remarkably well on previously unseen coding challenges, regardless of the degree to which it truly "understands" the task.

Yet it seems inherently disappointing if all the specialized knowledge about the structure of computer programs was ultimately of limited practical value for LLMs trained on lots of data. Intuitively, understanding the structure of programs should be important to building ML models that can write them. And maybe this will eventually be the case; there is room left for improvement in the CodeContest benchmarks (AlphaCode solves ~30% of the problems), and perhaps there will be future systems that leverage the power of AlphaCode combined with more classical structured program synthesis to improve results. But the possibility should be taken seriously that this is not true, that data and scale are all that is needed, once referred to as "the bitter lesson" (*8*). Thus, in addition to being a compelling achievement, AlphaCode is perhaps best viewed as a baseline for the power of "raw" large-scale models to write code. If "hybrid" ML methods that combine data-driven learning with engineered knowledge can perform better on this task, let them try. AlphaCode cast the die. The datasets are public. Let us see what the future holds. ■

[1]Computer Science Department, Carnegie Mellon University, Pittsburgh, PA, USA. [2]Bosch Center for AI, Pittsburgh, PA, USA. Email: zkolter@cs.cmu.edu

REFERENCES AND NOTES

1. Y. Li *et al.*, *Science* **378**, 1092 (2022).
2. T. Brown *et al.*, *Adv. Neur. Inf. Process Syst.* **33**, 1877 (2020).
3. M. Chen *et al.*, arXiv:2107.03374 [cs.LG] (2021).
4. E. Caballero, OpenAI, I. Sutskever, Description2Code dataset. GitHub (2021); https://bit.ly/3Ogx18I.
5. R. Pur *et al.*, arXiv:2105.12655.1035 [cs.SE] (2021).
6. S. Gulwani *et al.*, *Found. Trends Program. Lang.* **4**, 1 (2017).
7. E. M. Bender, T. Gebru, A. McMillan-Major, S. Shmitchell, in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (2021), pp. 610–623.
8. R. Sutton, "The bitter lesson." *Incomplete Ideas* (blog) 13 March 2019.

10.1126/science.add8258

POLICY FORUM

CLIMATE POLICY

Fairness considerations in global mitigation investments

Current mitigation finance flows are inadequate and unfair

By **Shonali Pachauri[1], Setu Pelz[1], Christoph Bertram[2], Silvie Kreibiehl[3], Narasimha D. Rao[1,4], Youba Sokona[5,6], Keywan Riahi[1]**

Despite overwhelming evidence that the world needs to make rapid and substantial investments in climate mitigation in this decade to meet the ambitious goals of the Paris Agreement (*1*–*3*), political and financial barriers continue to hinder mitigation efforts (*2*). Global mitigation investment pathways modeled in the sixth assessment report (AR6) of the Intergovernmental Panel on Climate Change (IPCC) reach global climate goals in a cost-effective manner. These are agnostic about who should finance these and how to fairly allocate costs and benefits of mitigation efforts. We apply equity considerations to global cost-effective mitigation investment needs and derive "fair-share" regional contributions, which describe the direction and magnitude of interregional financial flows that align with each consideration. We find that flows from North America and Europe to other regions would have to increase substantially relative to present levels to meet the Paris Agreement goals under most equity considerations.

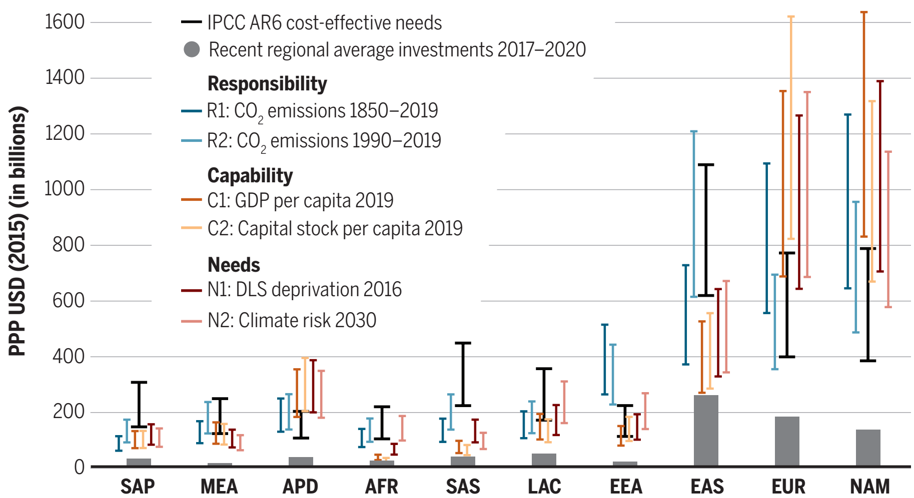
Progress on the alignment of financial flows with low greenhouse gas (GHG) emissions pathways remains slow (*3*). In 2019 and 2020, annual global climate finance flows were about USD (2015) 630 billion (with more than 90% for mitigation), but growth has slowed recently. The IPCC's AR6 stresses that these flows must increase globally by a factor between three and six to meet average annual needs until 2030 to avoid the most dangerous impacts of climate change. Adequate capital and liquidity for this is globally available, as is evident from the USD 2.4 trillion world energy investment in 2022 estimated by the International Energy Agency. The IPCC report also states that "accelerated financial international cooperation is a critical enabler of a low-greenhouse gas and just transition." In particular, adequate international support for near-term investment is essential to ensure that national policies are put in place to attract the required finance in this decade.

Global mitigation investment needs in the IPCC's AR6 are based on pathways generated by integrated assessment models (IAMs). Several recent critiques of IAMs engage with the history, assumptions, limitations, and normative positioning of such exercises (*4*). Although several studies propose fair global carbon budget–sharing schemes, few focus on equity considerations in the financing of mitigation investments, and these largely disregard near-term investment flows (*5*). We build on literature that argues that cost-effective mitigation investments require recognition of differentiated responsibilities, capabilities, and needs to yield an equitable outcome and be realized (*6*).

Consistent with suggestions in previous literature, we find that distributive justice considerations in global climate mitigation will require substantial interregional finance flows (*7*). Although mitigation activities involve costs that are distinct from investments, our work focuses specifically on modeled estimates of regional mitigation investment needs. This work provides a pathway to address the retrospective and prospective perspectives on climate equity in the literature: first, that wealthier high-emitting countries have historically benefited from fossil energy at the cost of poorer low-emitting countries (*8*), and second, that for cost-effective mitigation pathways to be fair, national and international redistributive measures are likely necessary (*9*). The magnitude and direction of interregional flows that we derive can also serve as input for policy and climate negotiations in the short term to ratchet up mitigation ambition, signal to the international private financial sector the magnitude of the required increase of interregional finance, and guide industrial pathways and value chain development toward a just and sustainable energy transition.

Regional needs and fair-share contributions

Annual regional cost-effective investment needs from 2020 to 2030 aligned with the well-below 2°C and 1.5°C compatible targets are shown (black, spanning the range from IPCC AR6). Recent regional average annual investments are shown for 2017 to 2020 (gray bars). Ranges of fair-share regional contributions to investment needs are shown using distinct considerations of equity (colors).



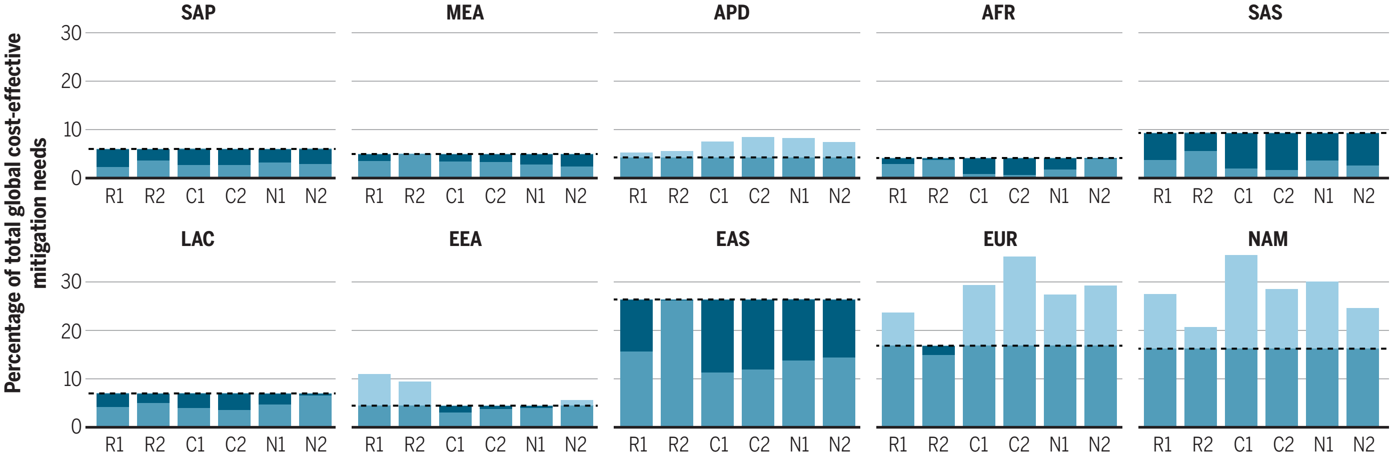
Regions are defined as follows: SAP, Southeast Asia and developing Pacific; MEA, Middle East; APD, Asia-Pacific developed; AFR, Africa; SAS, southern Asia; LAC, Latin America and Caribbean; EEA, Eastern Europe and west-central Asia; EAS, Eastern Asia; EUR, Europe; NAM, North America. DLS, decent living standards; GDP, gross domestic product; IPCC AR6, Intergovernmental Panel on Climate Change sixth assessment report; PPP, purchasing power parity

GRAPHIC: C. BICKEL/*SCIENCE*

[1]International Institute for Applied Systems Analysis, Laxenburg, Austria. [2]Potsdam Institute for Climate Impact Research, Potsdam, Germany. [3]Germanwatch e.V., Bonn, Germany. [4]Yale School of the Environment, New Haven, CT, USA. [5]South Centre, Geneva, Switzerland. [6]Department of Science, Technology, Engineering, and Public Policy, University College London, London, UK. Email: pachauri@iiasa.ac.at

Gaps between needs and fair-share contributions

Regional lower-bound cost-effective mitigation investment needs (dashed lines) and fair-share contributions under distinct considerations of equity (Rs, Cs, and Ns, where R is responsibility, C is capability, and N is needs) are shown as shares of total lower-bound global mitigation investment needs (see fig. S2 for upper-bound shares and fig. S3 for flows as share of regional gross domestic product). Some within-region needs can be met by fair-share contributions from countries in the region (●). Some regions' fair-share contributions exceed their needs and can help meet needs in other regions (●). Some regions' needs exceed their fair-share contributions, leaving gaps to be met by contributions from other regions (●).



Regions are defined as follows: SAP, Southeast Asia and developing Pacific; MEA, Middle East; APD, Asia-Pacific developed; AFR, Africa; SAS, southern Asia; LAC, Latin America and Caribbean; EEA, Eastern Europe and west-central Asia; EAS, Eastern Asia; EUR, Europe; NAM, North America. R1, CO_2 emissions 1850–2019; R2, CO_2 emissions 1990–2019; C1, gross domestic product per capita 2019; C2, capital stock per capita 2019; N1, decent living standards deprivation 2016; N2, climate risk 2030

EQUITY CONSIDERATIONS AND INDICATORS

We begin with the range of cost-effective regional mitigation investment needs in the decade 2020–2030 to achieve targets compatible with well below 2°C and 1.5°C warming, as provided in the IPCC AR6 (*3*). Regions are made up of countries and territories collected into broad geographical groups following the IPCC country grouping (see table S1). Regional mitigation investment needs [defined in supplementary materials (SM) section 2 and table S2], include low-carbon energy resource extraction, conversion, power generation, transmission, and storage, as well as economy-wide energy efficiency improvements (*3*).

Our subsequent consideration of equity is consistent with principles emerging in the climate equity literature that correspond to considerations of responsibility, capability, and needs, which in turn underlie the notion of "Common but Differentiated Responsibility and Respective Capabilities (CBDR-RC) in light of national circumstances" enshrined in the Paris Agreement (*6*). We apply these equity considerations to allocate "fair-share" regional contributions to annual, cost-effective, global climate mitigation investment needs in the decade 2020–2030 in proportion to the indicators we select for each (see tables S4 and S5).

We draw on the literature in selecting established indicators for regional responsibility and capability and introduce two new indicators that describe regional needs, which have not previously been operationalized in the IAM literature. We measure responsibility (R) as historical cumulative emissions shares. Given divergent views on when countries should be held accountable for their emissions, we implement two periods, one from 1850, accounting for postindustrial contributions, and the other since 1990, the year of the first IPCC assessment report. We consider only emissions of the dominant long-lived climate forcer, carbon dioxide (CO_2), from fossil fuel and industry, because other GHG emissions have not yet been thoroughly explored in the climate equity literature.

For capability (C), we use per capita gross domestic product (GDP) (C-1) and per capita capital stock (C-2), an alternative wealth indicator that reflects the extent of physical fixed assets and infrastructure in an economy. Under the consideration of needs (N), we use the average degree of deprivation across distinct dimensions of human well-being encompassed by the decent living standards (*10*) (N-1) and the modeled share of regional population facing multisector climate risk in the year 2030 (*11*) (N-2). We propose these needs indicators because they are not composites of the others used (such as per capita GDP) and capture both retrospective and prospective aspects of climate equity in terms of achieved human well-being and future vulnerability to climate effects. (See SM for a complete description of the methods used to transform each indicator into a corresponding regional fair-share contribution.)

GLOBAL SCALE FINANCE FLOWS

The IPCC AR6 reported large investment gaps between the recent average investment levels (2017–2020) and cost-effective investment needs over the decade 2020–2030. For most regions, it reported that recent investments were about three to four times lower in magnitude than the cost-effective needs (see the figure). However, in some regions, the gaps are much wider. We find that these gaps shift dramatically when principles of equity are considered to derive fair-share contributions, requiring large interregional financial flows (see the figure). Our estimated range of interregional flows to meet fair-share regional contributions is between purchasing power parity (PPP) USD (2015) 248 billion and PPP USD (2015) 1581 billion annually during 2020–2030 (see fig. S1).

The magnitude of interregional financial flows required is lowest (i.e., closest to cost-effective needs) when considering responsibility for historical cumulative CO_2 emissions since 1990 (R-2) and highest when considering the capabilities of regions (C-1 and C-2). With the exception of R-2, fair-share contributions under any equity consideration would be far higher than cost-effective needs in North America and Europe and lower in Africa, South and Southeast Asia, and Latin America. That is, regional cost-effective investment needs in low- and middle-income countries (LMICs) are higher than their fair-share contributions under most equity considerations (see the figure). Accounting for cumula-

GRAPHIC: C. BICKEL/*SCIENCE*

tive CO_2 emissions since 1990 favors North America and Europe over other regions that experienced much of their industrial growth in recent decades. By contrast, we see that capability- and needs-based allocations (C- and N-) require substantial [PPP USD (2015) 657 billion to PPP USD (2015) 1.581 trillion] mitigation finance flows to regions dominated by LMICs to bridge the gap between regional cost-effective investment needs and fair-share contributions. As a practical matter, and given differing notions of fairness, policy-makers may want to combine and weight equity considerations to find consensus on representing fair efforts in international negotiations. The adoption of more ambitious pledges in LMICs may depend on such consensus to guide stronger commitments to international financial flows. To aid this, we have developed a tool that allows for selection and weighting of equity considerations and corresponding indicators (https://data.ece.iiasa.ac.at/fairmitigation/).

"...we find that distributive justice considerations in global climate mitigation will require substantial interregional finance flows."

IMPLICATIONS OF FAIRNESS

Globally, climate mitigation investment gaps must be bridged to meet the agreed temperature goals of the Paris Agreement. How to finance global cost-effective climate mitigation investment needs across regions is still debated. We show that near-term interregional financial flows consistent with respecting equity principles enshrined in the Paris Agreement are substantial. Our proposal helps derive both the magnitude and directions of interregional flows that are necessary to incorporate selected considerations of equity, providing a way to integrate principles of equity into established approaches for developing global cost-effective mitigation scenarios. Importantly, our results indicate that interregional flows must be scaled up no matter which combination of our selected equity considerations and indicators we consider.

This work is consistent with other recent efforts to consider equity more explicitly in modeled mitigation pathways through alternative methods, such as applying specific social welfare functions, projecting degrowth in the Global North or more convergent growth futures, and explicitly accounting for regional populations in poverty to safeguard and exclude them from mitigation efforts in the near term. In reflecting on recent critiques of the AR6 IPCC mitigation pathways, we show here that cost-effective mitigation pathways are consistent with CBDR-RC as enshrined in the Paris Agreement when equity considerations guide the allocation of necessary financial flows. Future modeling efforts that explicitly represent the finance sector and specific economic instruments can also help derive fair interregional flows endogenously.

Our proposal and accompanying tool to derive fair-share near-term regional contributions to global mitigation investment needs does not yet address the much-needed investments to meet both regional climate adaptation needs and loss and damages, which are a priority for most LMICs (*12*). Moreover, we recognize that the current target of USD 100 billion per year promised by developed nations for climate action in developing nations under the Paris Agreement has been problematic. Among the key complexities of this target is that it serves, at times, two distinct purposes: both to redistribute resources from developed to developing nations and to mobilize the scale of finance needed to achieve the Paris Agreement targets (*13*). It also does not clearly differentiate between adaptation and mitigation needs.

Although our work exploring the magnitude and direction of interregional financial flows that must be mobilized under different considerations of equity can inform negotiations, agreement around new targets for mitigation and adaptation cost support as well as international redistribution will clearly need to deal with these issues. What is clear from our results is that even when considering responsibility for historical cumulative CO_2 emissions since 1990 (R-2), which is an indicator most favorable to regions of the Global North, the magnitude of annual interregional flows just for mitigation action alone must increase substantially to PPP USD (2015) 250 billion to PPP USD (2015) 570 billion in the near term.

Several caveats apply to our proposal. There are other justice considerations and intraregional distributional concerns that we do not consider, such as claims for committed climate impacts or how local benefits and returns can be fairly distributed. Interregional financial flows, as implied by our paper, can be mobilized through a number of different instruments, and each may have its own implications in terms of political feasibility and economic effectiveness. This is particularly important because regional investment risk profiles differ, which may also hinder financial flows. Finally, we do not address the wider issue of the distribution of macroeconomic costs and benefits resulting from the investments, which may change over time and require adaptive frameworks to motivate ambitious global mitigation strategies. A recent proposal toward fair-efforts metrics suggests one approach to account for true costs of mitigation technologies and nonmonetary benefits that may help advance this understanding (*14*).

Clear institutional and regulatory frameworks are needed to mobilize the magnitude of finance that is required to achieve globally agreed upon climate targets. Agreement on how to redirect international and domestic finance toward urgent near-term mitigation investments and climate adaptation efforts will be critical. Continued neglect of differentiated responsibilities, capabilities, and needs in the regional allocation of mitigation investment contributions risks lose-lose outcomes. Interregional cooperation will be necessary to move past this gridlock. Our work describes one pathway toward finding consensus by embedding distinct considerations of distributional justice in the derivation of "fair" regional contributions to globally cost-effective mitigation investment needs. This can inform estimates of the support required to bridge interregional financing gaps. Progress here will serve as a clear signal to governments, industry, and nongovernment actors and will be crucial for building the necessary momentum in regions where finance is scarce. ■

REFERENCES AND NOTES

1. D. Tong *et al.*, *Nature* **572**, 373 (2019).
2. H. D. Matthews, S. Wynes, *Science* **376**, 1404 (2022).
3. IPCC, *Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, P. R. Shukla *et al.*, Eds. (Cambridge Univ. Press, 2022).
4. J. T. S. Pedersen *et al.*, *Glob. Environ. Change* **75**, 102538 (2022).
5. M. Tavoni *et al.*, *Clim. Change Econ.* **4**, 1340009 (2013).
6. K. Dooley *et al.*, *Nat. Clim. Change* **11**, 300 (2021).
7. N. Bauer *et al.*, *Nature* **588**, 261 (2020).
8. C. W. Callahan, J. S. Mankin, *Clim. Change* **172**, 40 (2022).
9. B. Soergel *et al.*, *Nat. Commun.* **12**, 2342 (2021).
10. J. S. Kikstra, A. Mastrucci, J. Min, K. Riahi, N. D. Rao, *Environ. Res. Lett.* **16**, 095006 (2021).
11. E. Byers *et al.*, *Environ. Res. Lett.* **13**, 055012 (2018).
12. *Nature* **609**, 8 (2022).
13. W. P. Pauw *et al.*, *Clim. Policy* **22**, 1241 (2022).
14. N. D. Rao, *PLOS Clim.* **1**, e0000069 (2022).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adf0067

Published online 17 November 2022
10.1126/science.adf0067



Edited by **Jennifer Sills**

Data-based tools can prevent crowd crushes

The formation of mass crowds, in the absence of external danger, should not pose a threat to human lives. However, deadly crowd crush incidents, such as the tragedy in Seoul on 29 October (*1*), continue to occur. The Seoul incident was reminiscent of several similar recent catastrophes (*2*), all of which involved an unpredictable crowd size, a restrictive space with no evacuation routes, unrestricted entry, and the absence of real-time density monitoring. The science of crowds can help to prevent future crowd crush disasters.

The principles of crowd turbulence are well understood (*3*), and the connection to crowd density is clear. When a crowd of several thousands of humans reaches critical density levels of 8 to 10 persons/m^2, the crowd acts like a continuous body (or fluid). Individuals lose control of their movement, and any small turbulence could magnify and propagate through the crowd, potentially creating instability, falls, and injuries. In such a scenario, there is little that individuals can do once trapped. The description of such incidents as a crowd "stampede"—implying misbehavior on the part of the people—obscures the root causes and hinders prevention.

Solutions are available. Crowd management needs to be recognized as a specialty and should be outsourced to professionals within the industry who have the expertise and tools to conduct estimation of the expected crowd size, risk analysis, and—most important—real-time monitoring and control. In doing so, the role of computer-vision and artificial intelligence methods in real-time estimation of flow and density is critical (*4*). Although an element of unpredictability may always exist in terms of the exact crowd size (especially in the case of events with no ticket sales), these tools can provide the ability to observe early signs of overcrowding and take appropriate actions.

It is important that interventional measures are implemented before the critical density levels and crowd instability are reached. Two metrics are key in making such determinations: Crowd density in a restricted area gives insight into the current state of the crowd, and the crowd flow provides information about the input rate, and thereby, the rate at which density is increasing. These two metrics together can provide an estimate of the time that it may take for the crowd to reach critical density levels.

A range of methodologies developed by computer vision scientists apply artificial intelligence to infer these important parameters from real-time surveillance video footage. Real-time estimation of crowd sentiment, using indicators such as social media activity (*5*, *6*), could further inform real-time interventional decisions. Crowd crushes are preventable, and the required tools and knowledge do exist. They only need to be implemented.

Milad Haghani[1]* and Ruggiero Lovreglio[2]
[1]Research Centre for Integrated Transport Innovation, School of Civil and Environmental Engineering, The University of New South Wales, Sydney, NSW 2052, Australia. [2]School of Built Environment, Massey University, Auckland, New Zealand.
*Corresponding author.
Email: milad.haghani@unsw.edu.au

REFERENCES AND NOTES

1. E. Bagshaw, S. Na, "'Absolutely avoidable': How the Seoul crowd crush unfolded," *The Sydney Morning Herald* (2022).
2. D. Helbing, A. Johansson, H. Z. Al-Abideen, *Phys. Rev. E* **75**, 046109 (2007).
3. V. G. Ivancevic, D. J. Reid, *Nonlinear Dynam.* **68**, 285 (2012).
4. B. Yogameena, C. Nagananthini, *Int. J. Disaster Risk Reduct.* **22**, 95 (2017).
5. V. X. Gong, W. Daamen, A. Bozzon, S. P. Hoogendoorn, *Transport. Res. Rec.* **2673**, 836 (2019).
6. S. Wakamiya *et al.*, in *International Symposium on Web and Wireless Geographical Information Systems* (Springer, 2015), pp. 37–49.

10.1126/science.adf5949

Co-management of Chile's escaped farmed salmon

Salmon escapes from aquaculture farms taint Chile's reputation as a sustainable global provider of salmonid products (*1*). Massive salmon escapes from sea-based farms in Patagonia are frequent; escapes occur 1 to 16 times per year, and about 5 million salmon escaped between 2010 and 2020 (*2*). Escapes result in socio-environmental risks (*2*, *3*) and may increase the likelihood that salmon will become invasive in Patagonia (*2*, *4*). Adverse effects of aquaculture salmon on native biodiversity have been well documented, including the declines of native and threatened freshwater fish populations (*2*, *5*, *6*). Traditional livelihood fishing activities of local communities have also been disrupted by escaped salmon (*6*). There is a clear need to reduce these impacts on Patagonia, a valuable global natural reserve (*2*).

In the past, escapes were considered the private property of aquaculture companies. Legislation compelled them to recapture at least 10% of escaped fish and prohibited capture by other parties, including bycatch (*7*). However, the companies only recaptured approximately one-third of escapees (*8*), whereas artisanal fishers have been responsive, mobile, and efficient in capturing and selling escapees in informal markets (*6*, *8*), a practice that remains illegal (*7*).

On 25 October, Chile's congressional committee passed a bill, which will be rectified by Congress, deprivatizing aquaculture salmon and trout escapes in bycatch and implementing tougher sanctions for the companies responsible for the escapes (*7*). The approved bill includes a recommendation by scientists and conservationist entities to allow artisanal and recreational fishers to increase fishing pressure on escaped salmon and trout and indicates that guidelines for legally selling captured salmon will be defined in future bylaws (*7*).

Co-management of Chile's salmon escapes

PHOTO: MAIT JURIADO PHOTO/GETTY IMAGES

Analyzing and monitoring crowds can help to prevent deadly crowd crush incidents.

by aquaculture companies, recreational, and artisanal fishers could increase the recapture rate, reducing the escapees' negative socio-environmental impacts on Patagonia. Norway, the largest salmon producer in the world (*9*), passed legislation in 2015 requiring license holders for salmon aquaculture to manage and fund efforts to recapture as many escapees as possible (*10*). Escaped salmon that elude recapture are subsequently managed as wild salmon for harvest (*11*). These changes have resulted in a decrease in both the number of escaped fish and the proportion of recaptured salmon caught on Norwegian coasts and fjords using bag nets and bend nets (*11*, *12*).

Co-management could also benefit local communities, if escaped fish are safe to eat. Some escaped fish may have been treated with antibiotics or other chemicals. Therefore, testing for such contaminants should be mandatory before human consumption.

Although the approved bill is a step in the right direction for the management of salmon and trout escapes, legislators and stakeholders should work toward a holistic approach that includes a long-term monitoring program (*2*) and a complete characterization of the socio-ecological systems of Patagonia affected by these escape events. The recent period of profound political changes in Chile could allow for a paradigm shift in conservation strategies.

Guillermo Figueroa-Muñoz[1,2]*, Francisco Correa-Araneda[3], Beatriz Cid-Aguayo[2,4], Alfonso Henríquez[2,5], Loreto Arias[2], Ivan Arismendi[6], Daniel Gomez-Uchida[2,7]

[1]Department of Wildlife, Fisheries, and Conservation Biology, University of Maine, Orono, ME 04469, USA. [2]Nucleo Milenio de Salmónidos Invasores Australes, Concepcion, Chile. [3]Unidad de Cambio Climático y Medio Ambiente, Instituto Iberoamericano de Desarrollo Sostenible, Universidad Autónoma de Chile, Temuco, Chile. [4]Departamento de Sociología, Facultad de Ciencias Sociales, Universidad de Concepción, Concepcion, Chile. [5]Facultad Ciencias Jurídicas y Sociales, Universidad de Concepción, Concepcion, Chile. [6]Department of Fisheries, Wildlife, and Conservation Sciences, Oregon State University, Corvallis, OR 97331, USA. [7]Departamento de Zoologia, Facultad de Ciencias Naturales y Oceanográficas, Universidad de Concepcion, Concepcion, Chile.
*Corresponding author.
Email: guillermo.figueroa@maine.edu

REFERENCES AND NOTES

1. D. Gomez-Uchida *et al.*, *Science* **361**, 857 (2018).
2. D. Soto *et al.*, *Rev. Aquac.* 10.1111/raq.12711 (2022).
3. J. G. Navedo, L. Vargas-Chacoff, *Science* **372**, 695 (2021).
4. I. Arismendi *et al.*, *Rev. Fish Biol. Fish.* **24**, 919 (2014).
5. E. Habit *et al.*, *Global Ecol. Biogeogr.* **19**, 697 (2010).
6. D. Soto *et al.*, *Ecol. Appl.* **11**, 6 (2001).
7. Honorable Cámara de Diputadas y Diputados, Chile, "Modifica la ley N°18.892, General de Pesca y Acuicultura, en materia de prohibición de captura de especies salmonídeas provenientes de cultivos de acuicultura" (2022); https://www.camara.cl/legislacion/ProyectosDeLey/tramitacion.aspx?prmID=12089&prmBOLETIN=11571-21 [in Spanish].
8. "Plantean cuatro hipótesis sobre destino de salmones escapados desde centro Mowi Chile," Salmon Expert (2020); https://www.salmonexpert.cl/escape-estmago-invasal/plantean-cuatro-hiptesis-sobre-destino-de-salmones-escapados-desde-centro-mowi-chile/1345252 [in Spanish].
9. B. Hersoug, *Aquaculture* **545**, 737144 (2021).
10. Ministry of Industry and Fisheries, "Regulation on joint responsibility for fishing etc. of escaped farmed fish" (2015); https://lovdata.no/dokument/SF/forskrift/2015-02-05-89 [in Norwegian].
11. T. Aronsen, E. Ulvan, T. Næsje, P. Fiske, *Aquac. Environ. Interact.* **12**, 371 (2020).
12. H. M. Føre, T. Thorvaldsen, *Aquaculture* **532**, 736002 (2021).

10.1126/science.adf6211

Long-term solutions for China's heat and drought

Since July, China has experienced the most severe heatwave ever recorded in the world, compounded by the worst drought since record-keeping began in the country six decades ago (*1*). The areas hardest hit by the combined heatwave and drought fall along the Yangtze River basin, a populous and immense industrial economic belt. Impacts range from disruptions in global supply chains (*2*) to reductions in habitat for critically endangered species such as the Yangtze finless porpoise. Long-term anticipatory policies are needed to mitigate these far-reaching effects, in China and beyond, yet the current policy approach favors short-term solutions that are ill-equipped to tackle persistent climatic extremes in an era of climate catastrophe.

Poyang Lake, China's largest freshwater lake, located along the Middle Reaches of the Yangtze River, is illustrative of the devastation triggered by this summer's extreme weather. Suffering from more than 100 days of drought and counting, the lake has shrunk by nearly 90% (*3*). Numerous rare fishes, porpoises, and more than 700,000 migratory birds have been affected (*3*). The number of Yangtze finless porpoises, for example, was about 700 in 2021 (*4*), but few were found this October, mainly due to grounding and food shortages triggered by the drought (*5*). Poyang Lake has been listed as a wetland of international importance since 1992 (*6*) and provides key habitat for overwintering migratory birds, including white cranes and red-crowned cranes, yet the ongoing drought jeopardizes this critical habitat (*3*).

China's drought is extreme, but not exceptional. Since July, extreme heatwaves and persistent drought have swept across Europe, North America, and Asia (*7*). Climate change has led to a 20-fold increase in the likelihood of Northern Hemisphere droughts in 2022, with the frequency, spatial extent, and severity anticipated to continue to increase throughout the century (*8*). As drought and extreme weather events become increasingly endemic, policy responses must shift from short-term reactive strategies to long-term proactive measures.

Establishing a drought early monitoring and warning system that operates at timescales longer than 2 years and at regional rather than local levels is paramount. Better drought information and predictions would permit water regulation among different sectors and industries across larger geographies and timeframes. Drought impacts can be mitigated by harnessing the full potential of engineering projects, such as transferring water to lakes and reservoirs above Poyang Lake to avoid ecological catastrophe (*9*). Nature-based solutions, such as wetland restoration, will help mitigate long-term impacts to critically endangered species. Protecting, managing, and using low-lying, shallow lake areas can provide critical habitat for species during low-flow periods. Through a range of engineered and nature-based solutions, drought-stricken regions can develop long-term approaches to mitigating extreme weather events in the age of climate change.

Xiaona Guo[1], Annah Zhu[2], Qiang Li[1], Zilong Xia[3], Ruishan Chen[1]*

[1]Shanghai Jiaotong University, Shanghai 200240, China. [2]Wageningen University, 6706KN Wageningen, Netherlands. [3]Nanjing University, Nanjing 210023, China.
*Corresponding author. Email: rschen@sjtu.edu.cn

REFERENCES AND NOTES

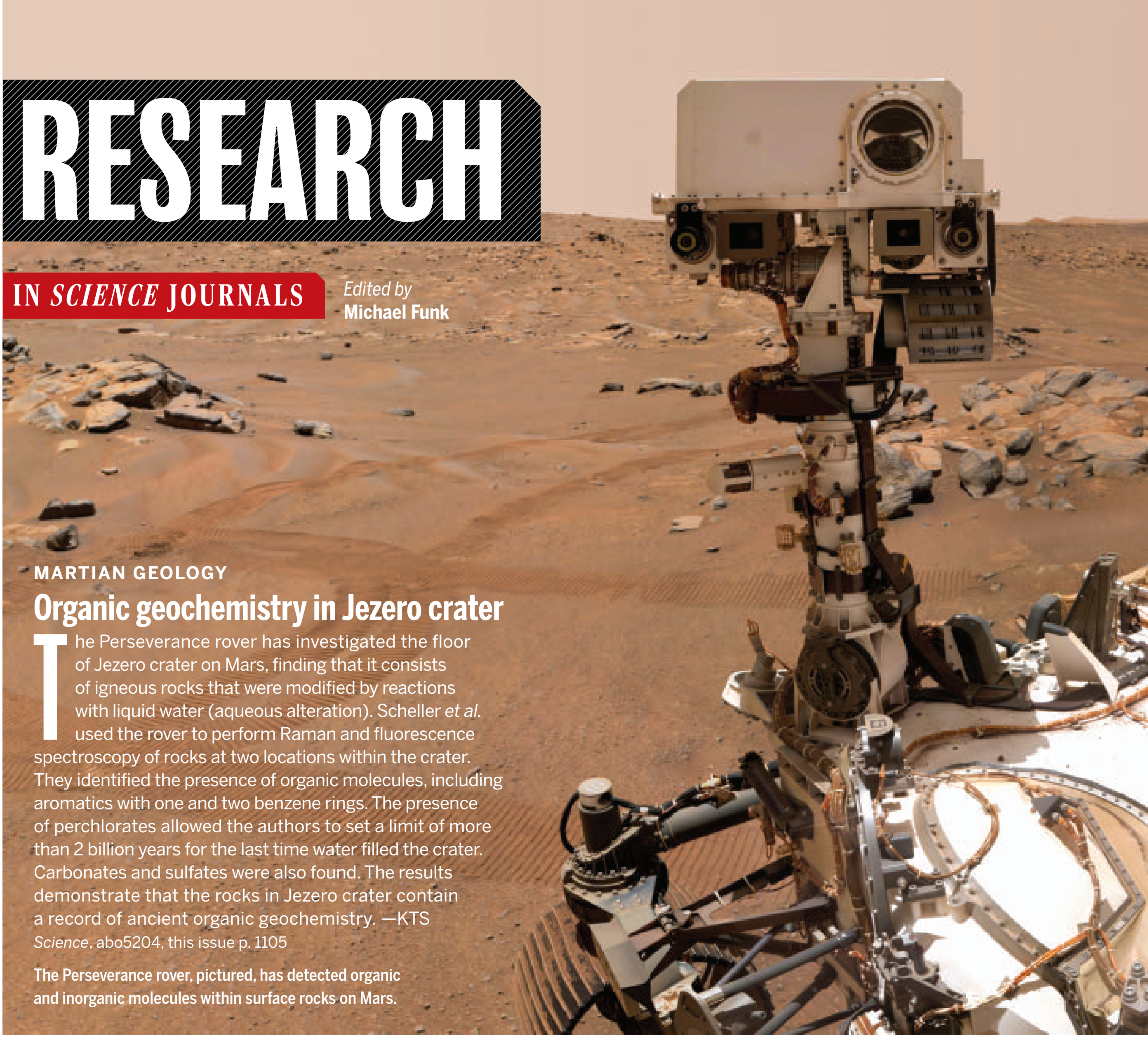
1. M. Page, "Heatwave in China is the most severe ever recorded in the world," *New Scientist* (2022).
2. R. Redondo, "Drought and alarming declines in river flows threaten global economy," *Atalayar* (2022).
3. J. Wang, W. Cang, "Biodiversity threatened in drought-hit Poyang Lake," *China Daily* (2022).
4. "The number of porpoises increased significantly after the water quality in Poyang Lake improved," *COP 15 News Center* (2022); https://www.cop15news.com/media/2022-08/24/c_1211678982.htm [in Chinese]
5. F. Lin, "First scientific survey on finless porpoises after the Yangtze River fishing ban, looking for porpoises in the arid Poyang Lake," South Weekend (2022); www.infzm.com/wap/#/content/236381 [in Chinese].
6. Y. Liu, G. Wu, X. Zhao, *Environ. Res. Lett.* **8**, 014010 (2013).
7. Z. Rosenthal, K. Patel, "Earth just experienced one of its warmest summers on record," *The Washington Post* (2022).
8. D. Schumacher *et al.*, "High temperatures exacerbated by climate change made 2022 Northern Hemisphere droughts more likely," *World Weather Attribution* (2022); https://www.worldweatherattribution.org/wp-content/uploads/WCE-NH-drought-scientific-report.pdf.
9. Y. Liu, "Poyang Lake: Caught between a dam and a sluice wall," *China Dialogue* (2021).

10.1126/science.adf6012

RESEARCH

IN *SCIENCE* JOURNALS

Edited by **Michael Funk**



MARTIAN GEOLOGY

Organic geochemistry in Jezero crater

The Perseverance rover has investigated the floor of Jezero crater on Mars, finding that it consists of igneous rocks that were modified by reactions with liquid water (aqueous alteration). Scheller *et al.* used the rover to perform Raman and fluorescence spectroscopy of rocks at two locations within the crater. They identified the presence of organic molecules, including aromatics with one and two benzene rings. The presence of perchlorates allowed the authors to set a limit of more than 2 billion years for the last time water filled the crater. Carbonates and sulfates were also found. The results demonstrate that the rocks in Jezero crater contain a record of ancient organic geochemistry. —KTS

Science, abo5204, this issue p. 1105

The Perseverance rover, pictured, has detected organic and inorganic molecules within surface rocks on Mars.

PNEUMONIA

A potent(iator) treatment for ARDS

Pneumonia can result in the development of acute respiratory distress syndrome (ARDS), which is characterized by increased permeability of the lung's endothelial barriers to fluid and protein. Targeted treatments for ARDS are limited. Erfinanda *et al.* noted downregulation of the cystic fibrosis transmembrane conductance regulator (CFTR) in pneumonia, resulting in increased endothelial permeability. The authors identified the mechanisms involved and showed that the CFTR potentiator ivacaftor stabilized CFTR abundance and rescued CFTR function in vitro. In wild-type mice, treatment with ivacaftor before or 6 hours after initiation of *Streptococcus pneumoniae* pneumonia resulted in improved barrier function and CFTR abundance and reduced signs of lung injury, suggesting a role for ivacaftor in the treatment of ARDS in humans. —MN

Sci. Transl. Med. **14**, eabg8577 (2022).

PHYSIOLOGY

Protection from PAH

Pulmonary arterial hypertension (PAH) is partially driven by pulmonary arterial vascular smooth muscle cell (PAVSMC) proliferation. Shen *et al.* investigated how PAH pathogenesis was regulated by TSC2, an inhibitor of cell growth–promoting signaling. TSC2 levels were reduced in pulmonary arteries and PAVSMCs from PAH patients, and TSC2 deficiency in smooth muscle in mice resulted in PAH. Restoring TSC2 abundance in rodent disease models improved lung function, suggesting that increasing TSC2 levels in PAVSMCs could reverse PAH, which currently lacks effective treatments. —WW

Sci. Signal. **15**, eabn2743 (2022).

GEOLOGY

A path for a high plateau

The timing of the uplift of the Qinghai-Tibet plateau influences how we interpret the atmospheric circulation and environment tens of millions of years ago. Miao *et al*. analyzed pollen records from a new sedimentary record to constrain the elevation of the northern part of the plateau. The authors found that the plateau likely obtained its current or higher elevation in the late Miocene, roughly 10 million years ago. —BG

Science, abo2475, this issue p. 1074

COMPUTER SCIENCE

AI masters Diplomacy

The game *Diplomacy* has been a major challenge for artificial intelligence (AI). Unlike other competitive games that AI has recently mastered, such as chess, Go, and poker, *Diplomacy*

PHOTO: NASA/JPL-CALTECH/MSSS

cannot be solved purely through self-play; it requires the development of an agent to understand other players' motivations and perspectives and to use natural language to negotiate complex shared plans. The Meta Fundamental AI Research Diplomacy Team (FAIR) *et al*. developed an agent that is able to play the full natural language form of the game and demonstrates performance well above the human average in an online *Diplomacy* league. The present work has far-reaching implications for the development of cooperative AI and language models for communication with people, even when interactions involve a mixture of aligned and competing interests. —YS

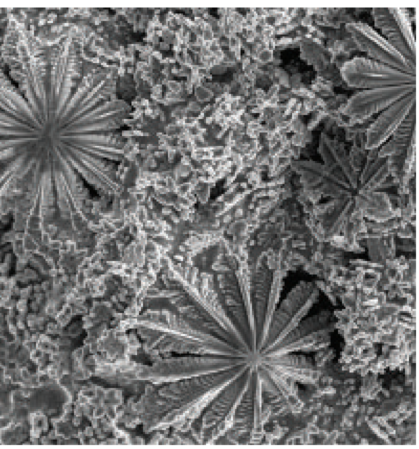
Science, ade9097, this issue p. 1067

CRYSTALLIZATION

Many metal morphologies

Metals will easily dissolve in liquid gallium at elevated temperatures and then precipitate out of solution when cooled down. However, the high surface tension of the liquid metal makes it very difficult to extract these crystals. Idrus-Saidi *et al*. developed an extraction method achieved by applying a voltage to the liquid metal solution while vacuum filtrating. The resulting crystals can have intricate morphologies, with some looking similar to snowflakes. This strategy should produce a variety of morphologies for metal particles across a wide range of different elements and alloys. —BG

Science, abm2731, this issue p. 1118



Scanning electron microscopy images of microscopic zinc crystals formed from a zinc–gallium alloy.

PHOTOS (LEFT TO RIGHT) IDRUS-SAIDI ET AL.; CHRISTIAN CHARISIUS/PICTURE ALLIANCE/GETTY IMAGES

ORGANIC CHEMISTRY

Biasing vinyl cation insertions

Forming a vinyl cation leaves a doubly bonded carbon center stranded with just six electrons instead of eight. Because of this deficiency, these intermediates tend to react extremely quickly and indiscriminately with any other bond in the vicinity. Nistanaki *et al*. report that a negatively charged, phosphorus-based catalyst can engage a vinyl cation just tightly enough to steer an ensuing intramolecular carbon–hydrogen insertion reaction toward just one of two mirror-image products. Isotopic labeling and computational analysis elucidate the likely mode of induction. —JSY

Science, ade5320, this issue p. 1085

ANTIBIOTIC RESISTANCE

Resilient in the long term

Tuberculosis caused by *Mycobacterium tuberculosis* is a persistent, sometimes lifelong infection that needs long courses of multiple antibiotics to treat. Not surprisingly, antibiotic resistance is rife. In a large sample of whole-genome data from clinical isolates, Liu *et al*. repeatedly observed an elongated phenotype showing rapid regrowth after typical regimens of antibiotic treatment. Signals of positive selection pointed to an essential transcriptional regulator (resR) and specific intergenic regions, which act in concert to regulate growth. Up to 10% of strains from high-tuberculosis-burden countries showed fixed mutations in these regions. Mutations across this cascade are associated with antibiotic treatment failure and are precursors to the emergence of classical antibiotic resistance. —CA

Science, abq2787, this issue p. 1111

IN OTHER JOURNALS

Edited by **Caroline Ash** *and* **Jesse Smith**



CLEAN ENERGY

Faster is better

Reducing greenhouse gas emissions by switching from fossil fuels to noncarbon energy sources is necessary to mitigate climate change, even though the transition will entail its own set of carbon dioxide emissions from energy and materials use. Lesk *et al.* show that the more quickly and aggressively we act to realize that transition, the better. They found that the energy needed to deploy renewable capacity sufficient for the more ambitious goal of limiting warming to 1.5°C is only 20% of that of a more gradual transition for a 2.0°C goal and only about 10% of that which will be needed by 2100 if current policies remain unchanged and warming is 2.7°C, as projected. —HJS

Proc. Natl. Acad. Sci. U.S.A. **119**, e2123486119 (2022).

Rapid deployment is key to minimizing the energy costs of the transition to a renewable energy future.

ECOSYSTEMS

All-seeing tiger shark's eye

Seagrass ecosystems are globally important because of their photosynthetic capacity to draw down and store carbon in the underlying sediment. Unfortunately, seagrass meadows are highly vulnerable to damage, and we only have a hazy idea of their true extent. Knowing that tiger sharks are faithful to seagrass ecosystems, Gallagher *et al.* attached cameras and/or pop-up satellite tags to 15 individual tiger sharks and exploited the animals' migratory behavior to monitor the extent of seagrass ecosystems around the Bahamas. Shark surveys were complemented by Landsat 8 images and human diver surveys, which also collected sediment cores for stored carbon estimation. Tiger sharks reached areas off-limits to human SCUBA divers, and seagrass could be observed growing at depth in these clear waters. The area covered by seagrass meadows was estimated to be 57,337 to



CONSERVATION

The insect apocalypse

Serious declines in insect numbers across species and systems have been reported. However, some studies show increases or equivocal results, leading to debate about what is really happening. Harvey *et al.* reviewed the effects of climate change, identifying the large number of threats that insects are currently facing and confirming that general patterns of decline are occurring. Overall, these threats fall into two categories: long-term responses to change and the acute impacts of extreme events. These threats are exacerbated by other anthropogenic factors, including habitat loss and agrochemical pollution. The authors conclude that overall damage to insect populations is high, and if the declines are not reversed, the impact on human activities will be significant if not catastrophic. —SNV *Ecol. Monogr.* 10.1002/ecm.1553 (2022).

Global insect decline is caused by a combination of factors, including climate change, pollution (such as pesticides, which killed the bees pictured), and habitat loss.

92,524 square kilometers, which roughly equates to 19.2 to 26.3% of the carbon buried in seagrass sediments worldwide. —CA
Nat. Commun. **13**, 6328 (2022).

IMMUNOLOGY

Old, particular, and exhausted

The immune system tends to become dysfunctional with age, resulting in increased chances of infection, autoimmunity, and cancer. This process, known as immunosenescence, is also associated with remodeling of lymphoid tissues. Ural *et al.* examined a cohort of human organ donors and found that lymph nodes associated with the lungs accumulated particulates derived from environmental pollutants as the individual aged. These particulates are preferentially taken up by a subset of $CD68^{+}CD169^{-}$ macrophages, which subsequently become functionally impaired and develop altered cytokine production. Moreover, these lymph nodes, which are strikingly black in color, display progressive anatomical disruption. The selective accumulation of air pollutants may be responsible, at least in part, for some of the immune defects experienced in late life. —STS
Nat. Med. 10.1038/s41591-022-02073-x (2022).

GENETIC RISK

Family history plus risk scores

With the increasing use of genome-wide association studies, interest has turned to polygenic risk scores (PRSs), which estimate the risk of genetic disease based on counts of disease-associated alleles in a genome. Given the costs of generating genetic data, there has been debate about whether PRSs are any better than family history at predicting disease risk—after all, family history captures both genetic and nongenetic factors. Mars *et al.* used data from more than 300,000 individuals in the FinnGen research project to compare the relative contributions of family history and PRS against disease risk. Across 24 diseases, the authors found that although these measures capture different information, their joint use improves disease prediction. The relative differences in predictive power of the two metrics differ across diseases, with implications for understanding how genetic architecture contributes to disease. —CNS
AJHG 10.1016/j.ajhg.2022.10.009 (2022).

QUANTUM OPTICS

Quantum frequency conversion

An optical-based quantum network connects individual nodes, transmitting information encoded in single photons between senders and receivers. However, because there are many different quantum platforms being developed that do not necessarily operate at same wavelength (even the nominally same platform can drift in wavelength), there is a need to convert the wavelength of the photons for compatibility. Bonsma-Fisher *et al.* demonstrated four-wave mixing in a photonic crystal fiber–based system that has the properties for efficiently converting photons over a broad range of wavelengths. Demonstrating the effect with communication-band photons at 1551 nanometers, the authors were able to convert single photons to any wavelength between 1226 and 1408 nanometers. Such wavelength conversion capability will be useful for linking up different quantum components across a network. —ISO
Phys. Rev. Lett. **129**, 2303603 (2022).

PHYSICS

A van der Waals quantum sensor

Solid-state defects, such as nitrogen vacancy centers in diamond, have been used as local sensors of magnetic field, pressure, and temperature. These sensors are typically embedded in three-dimensional structures, making it tricky to optimize the sensor-sample interface. Healey *et al.* instead used defects in hexagonal boron nitride (hBN), a van dear Waals material that can be easily exfoliated. After creating boron vacancy centers in thin layers of hBN by ion irradiation, the researchers interfaced these layers directly with the sample. The technique enabled them to simultaneously image magnetic field and temperature maps of the van der Waals magnet $CrTe_2$. The sensitivity and resolution of the method might be further improved by optimizing the materials and the thickness of the sensing layer. —JS
Nat. Phys. 10.1038/s41567-022-01815-5 (2022).

PHOTO: MAJA HITIJ/GETTY IMAGES

ALSO IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

MEDICINE

Clinical benefits of psychedelics

Classic psychedelic drugs, such as psilocybin and *N*,*N*-dimethyltryptamine (DMT), have been reported to have sustained benefit for various neuropsychiatric disorders and could have broader applications. What will it take for classic psychedelic drugs to be used in clinical settings? In a Perspective, Schindler and D'Souza discuss the evidence for the clinical use of psychedelics and their proposed mechanism of action to induce neuroplasticity. They also highlight the key issues to be overcome, such as understanding dosing and attendant safety questions and the problems with ensuring that clinical trial participants do not know they are taking the active drug. Nonetheless, psychedelics, either as single agents or as an assist, offer an opportunity to develop new therapeutic avenues for difficult-to-treat disorders, including addiction and depression. —GKA

Science, abn5486, this issue p. 1051

BATTERIES

Progress and challenges for electrolytes

Compared with the development of new cathodes and anodes, there has been less of a focus on the development of electrolytes. However, it is the electrolyte that controls the flow of ions and charges, and it is the only component in intimate contact with all the others. With the push toward higher energy and power densities, electrolytes are also involved in kinetically formed interphases that aid in the stability of a battery but can also hamper its operation. In a review, Meng *et al*. captures a number of trends that have emerged in the development of advanced battery electrolytes. —MSL

Science, abq3750, this issue p. 1065

SYSTEMS BIOLOGY

Doing the math on the central dogma

Gene expression can in theory be modulated at the level of transcription or translation, but both of these processes have constraints that complicate prediction of their outputs. To obtain a better quantitative understanding of the control of gene expression in bacteria, Balakrishnan *et al*. measured promotor on-rates, messenger RNA abundance, and protein abundance for more than 1500 genes in the bacterium *Escherichia coli* under many different growth conditions. Protein abundance largely reflects gene promoter on-rates and transcription, but has to comply with general constraints that keep the protein concentration constant and limit the number of ribosomes—and thus translational capacity. The authors propose a balancing of transcription with translation through Rsd, a factor that controls the availability of RNA polymerase. Their results may be useful in the design of synthetic circuits in bacteria and the prediction of their behavior in various growth conditions. —LBR

Science, abk2066, this issue p. 1066

EVOLUTIONARY ECOLOGY

Selection by modern agriculture

Intensive agriculture creates extreme environmental changes, including soil disturbance, water and nutrient addition, and application of chemical pesticides. Kreiner *et al*. documented how these extreme environments have driven changes to a native North American agricultural weed, common waterhemp (*Amaranthus tuberculatus*), using genomic data from paired agricultural and natural settings, as well as historic herbarium samples from both habitat types (see the Perspective by Waselkov and Olsen). The authors found that eastward range expansion of a distinct southwestern variety introduced new genetic diversity and facilitated rapid adaptation to agricultural environments in genes related to growth, environmental tolerance, and herbicide resistance. —BEL

Science, abo7293, this issue p. 1079; see also ade4615, p. 1053

COMPUTER SCIENCE

Machine learning systems can program, too

Computer programming competitions are popular tests among programmers that require critical thinking informed by experience and creating solutions to unforeseen problems, both of which are key aspects of human intelligence but challenging to mimic by machine learning models. Using self-supervised learning and an encoder-decoder transformer architecture, Li *et al*. developed AlphaCode, a deep-learning model that can achieve approximately human-level performance on the Codeforces platform, which regularly hosts these competitions and attracts numerous participants worldwide (see the Perspective by Kolter). The development of such coding platforms could have a huge impact on programmers' productivity. It may even change the culture of programming by shifting human work to formulating problems, with machine learning being the main one responsible for generating and executing codes. —YS

Science, abq1158, this issue p. 1092; see also add8258, p. 1056

CHEMICAL GENETICS

Beyond traditional drug design

Drug design prioritizes compounds likely to cross the cell membrane, leading to rules governing the size, polarity, and rigidity of chemical structures. However, there is increased interest in linking two chemotypes to allow greater selectivity for one target or to associate two targets. Lou *et al*. focused on a bitopic inhibitor, Rapalink-1, of the signaling protein mTOR that is active in vivo even though it is much too large to be considered drug like (see the Perspective by Lokey and Pye). The authors used chemical genetic approaches to show that the activity of Rapalink-1 depends on interferon-induced transmembrane proteins (IFITMs). They went on to show that IFITMs promote the cellular uptake of diverse linked chemotypes and to explore the role of linker length on the uptake pathway. This work may guide the development of bitopic drugs. —VV

Science, abl5829, this issue p. 1097; see also adf4412, p. 1054

PIEZOELECTRICS

Get the lead out

Piezoelectric materials convert an electrical signal into mechanical strain, making them vital as actuators and sensors. However, the best materials currently in use contain lead. Huangfu *et al*. found that doping a potassium–sodium niobium oxide with strontium improved the piezoelectric properties. This material had a high strain response at electric fields similar to its leaded counterparts and could potentially be a replacement for some lead-based piezoelectrics. —BG

Science, ade2964, this issue p. 1125

VACCINES

Boosting up flu-specific lung T cells

Despite the rapid embrace of messenger RNA (mRNA) vaccine technologies catalyzed by the COVID-19 pandemic, the extent of the resident memory T cell (TRM) response in the lung after different routes of

mRNA vaccination remains uncertain. By immunizing mice with a self-amplifying mRNA vaccine for influenza nucleoprotein encapsulated in dendritic nanoparticles, Künzli *et al.* assessed how various routes of vaccine administration affected CD4 and CD8 TRM responses. Whereas intramuscular prime-boost immunizations were sufficient to induce respiratory TRM, an additional intranasal boost further expanded both circulating and lung-resident CD4 and CD8 TRMs. These findings indicate that optimization of the routes of administration and sequence of mRNA vaccine doses can elicit robust antibody and lung TRM responses to pathogenic respiratory viruses. —IRW

Sci. Immunol. **7**, eadd3075 (2022).

REVIEW SUMMARY

BATTERIES

Designing better electrolytes

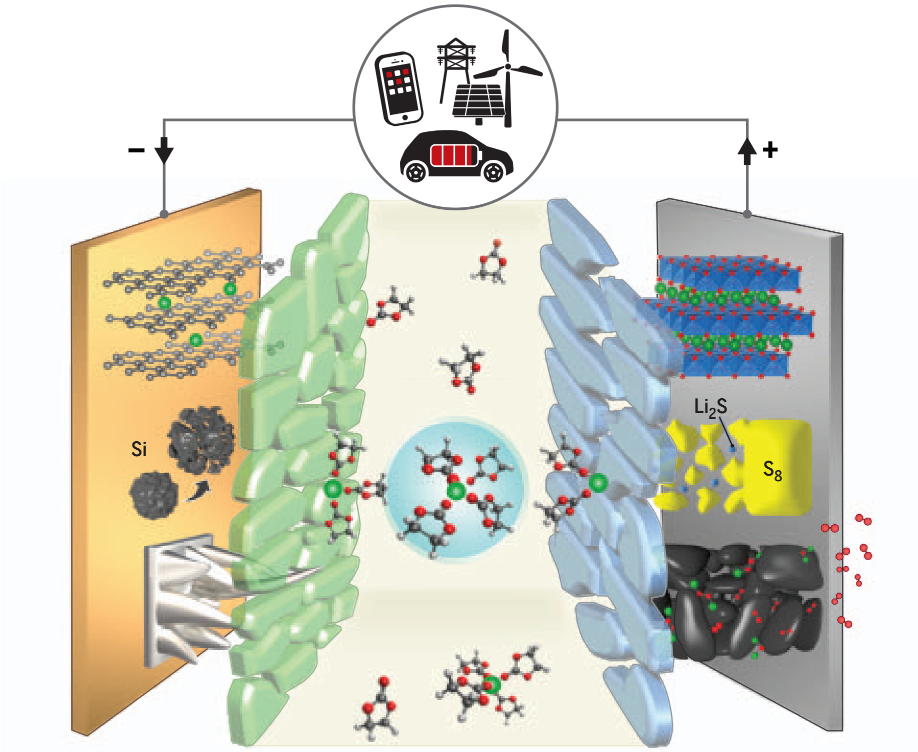
Y. Shirley Meng*, **Venkat Srinivasan***, **Kang Xu***

BACKGROUND: The electrolyte is an indispensable component in every electrochemical device, including lithium-ion batteries (LIBs). It physically segregates two electrodes from direct electron transfer while allowing working ions to transport both charges and masses across the cell so that the cell reactions can proceed sustainably. Whether powering our phones, driving our vehicles, or harvesting the intermittent energy from solar and wind farms, electrolytes in these LIBs determine how fast and how many times our devices can be recharged or how efficiently energy can be captured and stored over the grid. Occasionally, when an LIB is pushed away from the designed electrochemistry pathways by various factors such as excessive heat, mechanical mutilation, or internal short circuits induced under extreme charging conditions, electrolytes are also responsible for the fire and explosion accidents that we read about in the news.

The electrolyte is the most unique component in a battery. Because it must physically interface with every other component, it is obligated to satisfy numerous constraints simultaneously, including rapidly transporting ions and masses, effectively insulating electrons, and maintaining stability toward the strongly oxidative cathode and strongly reductive anode. Historically, the electrolyte-anode interfacing was the last piece of the puzzle to complete modern LIB chemistry.

ADVANCES: The commercial success of LIBs has attracted intense interest and investments in electrolyte research, which led to the identification of interphases as the key component responsible for the stable and reversible operations of cathode and anode materials far beyond the thermodynamic stability limits of any known electrolyte. These interphases, often with nanometer thickness, are formed by electrolytes in a self-limiting decomposition process, and they ensure fast rates of charging and discharging, maximum voltage, and reversibility of LIBs. In the past three decades, the chemistry, morphology, and formation mechanisms of interphases have been thoroughly investigated. Researchers have learned how such interphases are structured and what key ingredients they comprise and, most importantly, how to tailor them using electrolyte engineering. Today, it is widely accepted that designing better electrolytes also implies designing the associated interphases for the electrode materials. Although the accurate prediction of interphasial chemistry remains difficult, and key fundamental properties of interphases such as the rate and mechanism of ion transport across interphases are still unknown, the structure of the ion solvation sheath has been identified as an effective tool that directs the formation process of interphases. Such knowledge has been driving a series of new electrolyte concepts for emerging battery chemistries.

OUTLOOK: Efforts are being made to develop battery chemistries that promise high energy density, rapid charging, low cost, high sustainability, and independence from elements or materials of high geopolitical or ethical risks. Each individual chemistry may demand a unique electrolyte and corresponding interphase, but a few universal trends emerge: (i) a superconcentration of salts is used to leverage unusual properties arising from the altered ion-solvation structures; (ii) both polymeric and inorganic materials are used to solidify electrolytes so that the aggressive lithium-metal anode can be harnessed with higher safety; (iii) efforts are made to identify the most effective interphasial ingredients so that an interphase of singular composition can be designed and artificially applied; (iv) liquefied gaseous components are used to expand the low-temperature limits of conventional electrolytes; and (v) unusual electrochemical behaviors are explored by confining ion-solvation sheaths in nano- or sub-nano environments. ■



Electrolytes and the associated interphases play the central role in supporting diversified battery chemistries. On the anode side (left), the electrolyte must form an interphase that prevents graphitic anode from exfoliation, tolerates the drastic volume changes of a silicon electrode, and suppresses the growth of a dendritic form of lithium metal. On the cathode side (right), an interphase is critical in preventing the irreversible reactions with electrolytes, maintaining the lattice structure of transition metal oxides, suppressing the cross-cell shuttling of polysulfide species, and assisting the complicated triphasial reactions of an air-cathode. In all of these scenarios, interphases must enable ionic transport while insulating electronic transport.

The list of author affiliations is available in the full article online.
*Email: conrad.k.xu.civ@army.mil (K.X.); vsrinivasan@anl.gov (V.S.); shirleymeng@anl.gov (Y.S.M.)
Cite this article as Y. S. Meng *et al.*, *Science* **378**, eabq3750 (2022). DOI: 10.1126/science.abq3750

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abq3750

REVIEW

BATTERIES

Designing better electrolytes

Y. Shirley Meng[1,2]*, Venkat Srinivasan[2,3]*, Kang Xu[3,4]*

Electrolytes and the associated interphases constitute the critical components to support the emerging battery chemistries that promise tantalizing energy but involve drastic phase and structure complications. Designing better electrolytes and interphases holds the key to the success of these batteries. As the only component that interfaces with every other component in the device, an electrolyte must satisfy multiple criteria simultaneously. These include transporting ions while insulating electrons between the electrodes and maintaining stability against electrodes of extreme chemical natures: the strongly oxidative cathode and the strongly reductive anode. In most advanced batteries, the two electrodes operate at potentials far beyond the thermodynamic stability limits of electrolytes, so the stability therein has to be realized kinetically through an interphase formed from the sacrificial reactions between electrolyte and electrodes.

All electrochemical energy devices (fuel cells, batteries, and supercapacitors) operate by separating charges (*1*), and the electrolyte is the indispensable component to enforce such charge separation (*2*, *3*). The electrolyte physically segregates two electrodes from direct electron transfer while allowing working ions to transport both charge and mass across the cell so that the cell's reactions can proceed sustainably. In a rechargeable battery, a qualified electrolyte not only conducts ions while insulating electrons (a property combination known as "electrolyte nature"), but also must remain inert to the cell reactions. The electrolyte dictates how fast the cell reaction can proceed (power density) and how many times a battery can be charged and discharged (reversibility). Although an electrolyte does not directly determine the energy of a battery (which is instead dictated by both electrodes), it does so indirectly by allowing or disallowing an electrode depending on whether electrochemical stability can be achieved between it and the electrolyte.

To maximize energy density, the electrodes in advanced batteries are often pushed to operate at extreme potentials, where no known electrolyte is thermodynamically stable. In advanced batteries such as lithium-ion batteries (LIBs), the stability between the electrolyte and electrodes (and consequently the reversibility of the battery) must be achieved kinetically. Thus, trace amounts of electrolytes are sacrificially decomposed, and their solid products are deposited on the electrode surfaces, forming a protection against further decomposition. This protection layer in batteries is called "interphase" (*4*). Interphase formed in this self-limiting manner is almost always identified on anode surfaces in LIBs, where it is referred to as a solid electrolyte interphase (SEI). When interphase occurs on a cathode surface, it is known as a cathode electrolyte interphase (CEI) (*5*), which is less definitive and complicated by several factors (*3*). First, most transition-metal oxide cathode materials already carry a native passivation layer of Li_2CO_3, which in turn reacts with acidic electrolytes in the battery environment through pure chemical pathways. Second, further evolution occurs through electrochemical pathway, leading to eventual CEI chemistry. Finally, the surface of the cathode material might also be involved in the evolution, creating an intermediate phase and thus smearing the clear definition of where CEI starts. Nevertheless, the electrolytes and interphasial chemistries are closely associated. Because the electrolyte is the only component in a battery that is in contact with every other component, designing better electrolytes implies tailoring and balancing a host of properties, ranging from bulk (e.g., ion solvation and transport and extended liquid structure) to interfacial structure and stability (e.g., preferential assembly and orientation of ions and molecules at Helmholtz planes and corrosion suppression) to interphasial chemistry and morphology. Although the irreversible reactions leading to interphase formation mainly occur during the initial activation cycles of batteries, the interphases thus formed will dictate the power, energy, and reversibility of the battery during its entire service life.

Differing from the two-dimensional interface in classical electrochemistry, interphase constitutes an independent, three-dimensional existence with special chemistry and morphology. With nanometer thickness (~2 to 20 nm) and extremely high sensitivity, interphases have been the least understood component in rechargeable batteries until recently, when various advanced characterization and computational techniques were developed to shed light on this critical component (*6*).

LIBs enabled by interphases

The energy density and rechargeability of LIBs rely on interphases (Fig. 1) (*3*). Because the potential of a graphitic anode in a LIB (~ 0.01 V versus Li) resides far beyond the thermodynamic stability window of the electrolyte, as marked by the highest occupied molecular orbitals (HOMOs) and lowest unoccupied molecular orbitals (LUMOs), the reversible lithiation/de-lithiation chemistry relies on the protection from interphases (*7*). The electrolytes used in state-of-the-art LIBs vary with manufacturers and the cathode chemistry used; however, they almost exclusively follow a skeleton composition consisting of lithium hexafluorophosphate ($LiPF_6$) dissolved in a mixture of carbonate ester solvents, which is the result of balancing the requirements from every battery component (*8*), especially the graphitic anode at low potential (~0.01 V versus Li^+/Li) and transition-metal oxides or phosphates cathodes at high potential (~3.5 to 4.5 V versus Li^+/Li). It is now widely accepted that the SEI in LIBs mainly comes from the reductive decomposition of carbonate solvents and salt anions. These species are usually arranged in stratified structures, with the innermost layer being more inorganic (LiF and Li_2O) and the outer layer more organic (semicarbonates, oxalates, alkoxides, and polymers). Of special interest is the unbalanced contribution to SEI chemistry from one particular solvent, ethylene carbonate (EC) (*9*).

On the basis of numerous ex situ or in situ/in operando characterizations, the solvation sheath structure of Li^+ has been identified as being a key factor in SEI chemistry (*9*–*11*). In the very first cycle, as the negatively charged anode attracts the solvated Li^+ to its surface, it is the solvent molecules in the solvation sheath that initially experience reductive decomposition, thus becoming the precursors of the nascent SEI. EC happens to be the solvent that preferentially solvates Li^+. It is such solvation-interphase correlation that makes EC a preferred SEI contributor (*9*, *12*, *13*). Although the semicarbonate produced by EC decomposition constitutes the main SEI ingredient, recent studies have found that the nascent interphase can be electrochemically oxidizable (*14*), whereas long-term cycling turns it into monocarbonate (*15*).

The capability of EC in strongly solvating Li^+ and in forming a protective interphase on a graphitic anode make it almost an indispensable

[1]Pritzker School of Molecular Engineering, The University of Chicago, Chicago, IL 60637, USA. [2]Argonne Collaborative Center for Energy Storage Science (ACCESS), Argonne National Laboratory, Lemont, IL 60439, USA. [3]Joint Center for Energy Storage Research (JCESR), Argonne National Laboratory, Lemont, IL 60439, USA. [4]Battery Science Branch, Energy Science Division, US Army Combat Capabilities Development Command (CCDC) Research Laboratory, Adelphi Laboratory Center, MD 20783, USA.
*Email: conrad.k.xu.civ@army.mil (K.X.); vsrinivasan@anl.gov (V.S.); shirleymeng@anl.gov (Y.S.M.)

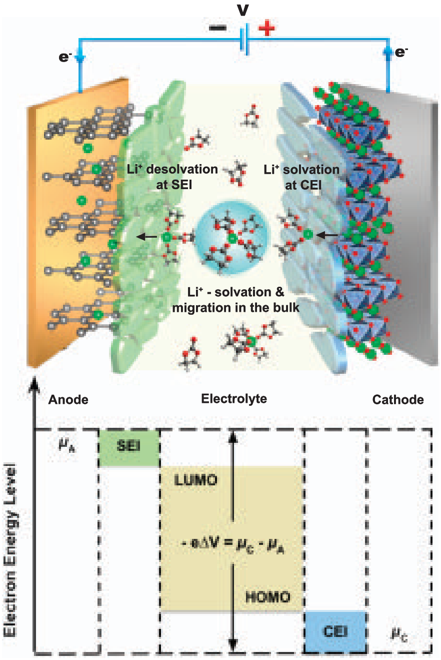


Fig. 1. How electrolytes work. The electrolyte must conduct ions, insulate electrons, and remain stable while simultaneously interfacing with all battery components. Substantial challenges arise when the cathode and anode operate at potentials beyond the region enclosed between LUMOs and HOMOs, where interphases must be formed to ensure the reversibility of the cell chemistry.

electrolyte component in all LIBs. However, the prevalent presence of EC in electrolytes also causes poor low-temperature performance and sluggish charging rates of LIBs because of its high melting point (36.5°C), high viscosity (2.0 cP at 25°C), and strong coordination with Li^+. Efforts have been made in recent years to replace EC with EC-free electrolytes based on acyclic carbonate such as ethylemethyl carbonate (melting point –53°C and viscosity 0.65 cP at 25°C) as the main solvent. When used in presence of fluoroethylene carbonate as a SEI-forming additive, this EC-free electrolyte could maintain most of the merits of EC-based electrolytes while circumventing the disadvantages (*16*, *17*).

Compared with SEI, there has been much less understanding about CEI. In certain cases, even the existence of CEI has been debated (*18*). Although studies show that conventional electrolytes could support cell operations up to 5.6 V under certain conditions (*19*), in cells that must withstand long-term cycling, the need for a CEI is still fundamentally driven by the mismatch of thermodynamic factors (Fig. 1). Unlike SEIs, which fully cover an anode, the depositions on the cathode surface, mostly inorganic fluorides and oxides, have often been found to be scattered and discontinuous, raising suspicion about whether they indeed function as protection (*20*). This was partially because the operating potentials of most cathode materials used in LIBs (<4.2 V) do not deviate too far from the limits allowed by electrolyte HOMOs (~4.0 V), but also because CEI chemistry is much more complicated. It involves not only electrolyte but also the decomposition and phase transformation of the cathode itself, thus causing the problem of defining where the interphase should start and end. However, with the development and commercial adoption of various high-voltage (>4.5 V) cathode materials in LIBs, the necessity for CEI is no longer questionable, just like SEI (*21*).

Researchers have reached a consensus that electrolyte and interphase are convoluted, and that the solvation structure for Li^+ (or any other cations essential to the cell chemistry, such as Na^+, Mg^{2+}, Ca^{2+}, and Zn^{2+}) is of vital importance when designing new electrolyte systems (*22*). The most direct fruition of the solvation-interphase correlation is perhaps the superconcentration concepts that led to diversified new electrolyte systems, including water-in-salt electrolytes (WiSEs) (*23*–*25*).

Designing electrolytes and interphases for emerging battery chemistries

In an era where LIB has become a mature technology, emerging battery chemistries promising either extremely high energy or power or low cost and high sustainability are being actively pursued. Alternative working ions such as Na^+ (*26*), K^+ (*27*), Mg^{2+} (*28*), Ca^{2+} (*29*), and even Al^{3+} (*30*) are also under consideration. To support these more energetic and much more aggressive chemistries that either operate at extreme potentials with drastic morphological and volume changes (Li^0 and Si), or involve extremely reactive surfaces (Li^0 and air cathodes), or rely on working ions difficult to move (Mg^{2+}, Ca^{2+}, and Zn^{2+}), the designing of electrolytes and the associated interphases must address different challenges.

On the anode side, Si is gradually replacing graphite as anode host for Li^+, which theoretically could provide 10 times higher specific capacity than graphite. However, the manyfold volume change of Si accompanying such high capacity presents severe challenges to the electrolyte and the associated SEI, which must withstand not only the low potential of Si (~0.5 V vs. Li^0) but also the mechanical stress and incessant exposure to new Si surfaces (*31*). Although making Si nanosized or embedding it in carbonaceous matrices could relieve the stress and fragmentation (*32*), fluorinated solvents such as fluoroethylene carbonate and the highly fluorinated salt anion fluorosulfonyl imide have also been found to stabilize Si effectively, leading to highly fluorinated SEIs containing LiF and organic fluorides that help to prevent Si fragmentation (*33*).

Much more severe challenges are associated with lithium-metal (Li^0) anodes, which are being revisited as the ultimate anode material because of their energy density (*34*). The promised high energy of Li^0 is accompanied by extreme reactivity, which is associated with excessive and incessant interphase growth, leading to both low efficiency of the Li^0 electrode and dangerous morphologies such as dendritic and dead Li^0. Identifying an electrolyte with effective SEI chemistry has become a heavily invested research direction, with the goal that Li^0 deposition could proceed homogeneously at high Coulombic efficiency (CE%). It is of special importance to circumvent fractal-like, long-aspect-ratio crystals (Fig. 2), which could either penetrate the cell and create an internal short circuit or be easily cut off near the roots and form highly dangerous dead Li^0 (*35*). The best electrolytes identified so far are based on fluorinated ether molecules rather than esters, which can achieve CE% as high as 99.9% and suppress dangerous Li^0 morphologies during long cycling (*36*). The improved Li^0 reversibility from such general fluorination of both electrolyte composition and interphasial chemistry has been attributed to the high surface energy for Li^0 to nucleate on the LiF.

Metals corresponding to multivalent cations (Mg^{2+}, Ca^{2+}, Zn^{2+}, and even Al^{3+}) are of high interest as possible anode materials with the

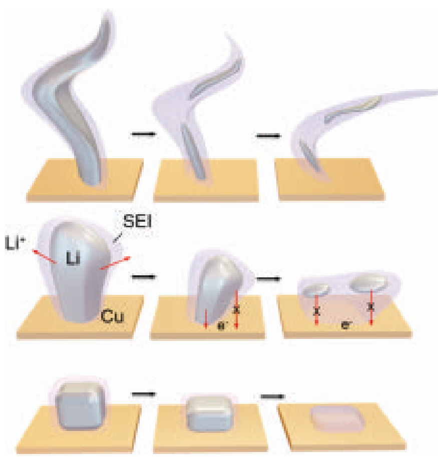


Fig. 2. The challenge of the Li^0 anode. Designing an electrolyte and the associated SEI to eliminate dangerous morphologies of dendritic and dead Li^0 is the key to enabling rechargeable lithium-metal batteries. Top: an inhomogeneous SEI encourages the growth of dendritic Li^0 at spots that are more conductive to Li^+, whereas dissolution near the root of the dendrite results in dead Li^0 that is electrically isolated from the current collector. Middle and bottom: designing better electrolytes and interphases guides the growth of dense Li^0 so that crystals of low aspect ratios are preferred, which means that they stay electrically attached during repeated plating and stripping.

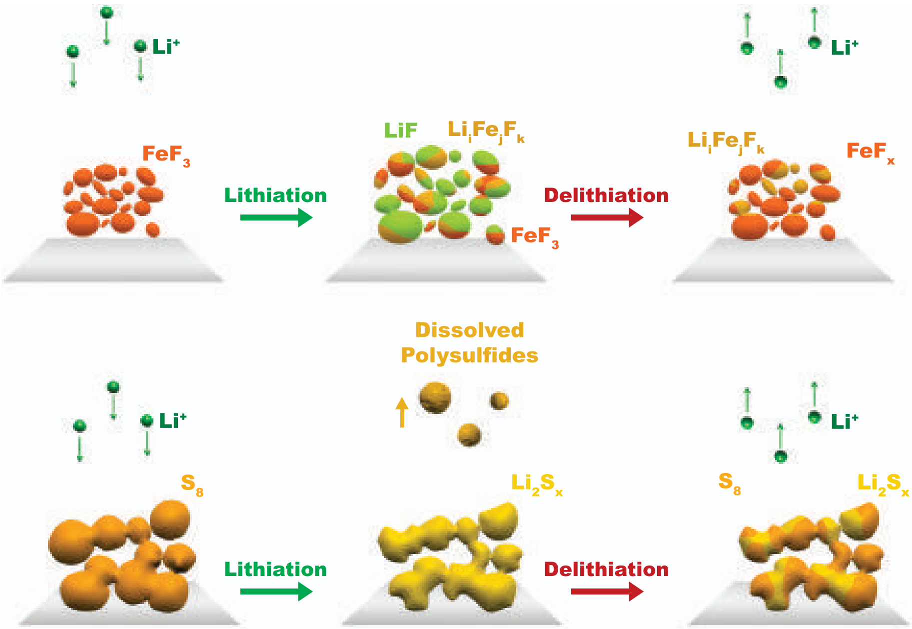


Fig. 3. Emerging conversion-reaction battery chemistries. Both iron fluorides (top) and sulfur (bottom) rely on drastic structure breaking and structure reforming to deliver high capacities and energy densities, but heir reversibility encounters severe difficulty because it is almost impossible to completely restore the original structure, and the active materials generated could be soluble in electrolytes and leave electrode bulk.

expectation that multiple electrons involved in their reactions would favor the specific capacity (*28*~*30*, *37*). However, the Coulombic force exerted by these multivalent cations on their environments (counterions, solvent molecules in bulk electrolyte, polyanions in lattice structure, or coordination sites in interphases) is so intense that either the multivalent salts have limited solubility in the electrolytes or the solvent molecules are so strongly bound to the multivalent ion that there is the possibility of their being electrostatically decomposed (*38*). Thus, desolvation at interphases becomes impossible or the multivalent cations are not able to travel across the interphases at all (*37*, *39*). Electrolyte design must resolve these problems, with the aim of forming (i) weakly coordinated electrolytes so that the cations are less solvated and the desolvation energy barrier could be more easily overcome, (ii) reduction-resistant electrolytes using various hydrides and organometallic Grignard reagents so that no SEI forms, or (iii) superconcentrated electrolytes in which networked anion clusters could assist in liberating cations for movement. Unlike LIB electrolytes, ethers rather than esters are exclusively favored here because of the intrinsic reactivity of the latter with these metals.

However, regardless of whether they are based on hydrides, Grignard reagents, or ethers, these reduction-resistant electrolytes raise new issues on the cathode side because they are all vulnerable against oxidation. Therefore, it remains to be seen whether these electrolytes can withstand the high voltage of multivalent cathode materials, which are still under development themselves. Prior experience with LIBs teaches us that the best approach to resolving such convoluted interests from the anode, cathode, and bulk electrolyte is an interphase that can effectively decouple them (*40*).

Among all multivalent cation chemistries that still largely remain remote from practical applications, Zn stands out as an exception. As the anode material in Alessandro Volta's first battery in 1799, Zn chemistries have powered us from the Industrial Revolution into the modern era, and they are still being used in the alkaline batteries available from drugstores. Zn reduction/oxidation potential sits at a moderate value (−0.76 V versus standard hydrogen potential) that allows for aqueous electrolytes; however, its poor reversibility has rendered these batteries essentially nonrechargeable. This situation has witnessed improvements in the past 5 years (*41*), when Zn^0 reversibility has progressed toward practical application, aided by advances in understanding Li^0 anode reversibility (*42*). The missing piece of a rechargeable Zn battery is a viable low-cost cathode material that could provide high reversibility and high capacity with Zn. The most promising candidates include air cathodes that use oxygen from the ambient air (*43*) and various intercalation hosts for Zn^{2+} (*44*).

On the cathode side, the pursuit for emerging chemistries has been focused on the search for either extremely high energy densities (*45*) or materials that are independent from Co or Ni, the two elements that LIB chemistries have been heavily reliant on, but are either extremely rare in the Earth's crust or are only available at high geopolitical or ethical risk.

As near-term goals, relatively more mature materials, such as improved versions of intercalation compounds, have been explored, some of which are already being gradually adopted by the LIB industry. Conceptually, they do not differ appreciably from the state-of-the-art LIB cathode materials, but the structural modifications of their lattices allow them to accommodate more Li^+ at higher potentials. The most popular approach to tailoring electrolytes for these aggressive cathodes are various electrolyte additives designed to have higher HOMOs so that even at a minor presence (from a parts-per-million level to a few percent) in the electrolyte, their oxidative decomposition could still dominate the CEI chemistry and form a better protection for both bulk electrolyte and the cathode itself (*46*). In some cases, the cathode structure is designed so that the redox reaction is no longer confined to the transition metals. Instead, the oxygen in the lattice could be harnessed for additional capacities. The electrolytes for such aggressive materials must display special stability against the reactive peroxide or superoxide species thus produced (*47*).

More radical materials leveraging conversion-reaction chemistries are being considered as long-term targets (*48*), including metal fluorides, sulfides (*49*), elemental sulfur (*50*), or even oxygen (*51*), which had been unsuccessfully explored in the early days of the rechargeable battery research. Unlike intercalation-type electrodes with minimal structural change, these electrode materials experience complete restructuring during the discharge, enabling significantly higher capacities, but at the expense of reversibility because it is almost impossible to completely restore the original electrode structures (Fig. 3). Because of advances made in the synthesis of nanostructured materials, these once-inaccessible chemistries could be revisited with higher feasibilities. The corresponding electrolytes and interphases are thus required to resist the strong catalytic activity from the nanosized metallic particles generated from the metal fluorides or sulfides, suppress the parasitic shuttling of soluble polysulfides across the cell, and remain inert to the reactive peroxide or superoxide intermediates from oxygen reduction (*48*–*51*). There have been limited improvements made thus far in addressing these challenges, leaving a wide space for researchers to explore.

New concept electrolytes and interphases

Substantial efforts are being made not only in designing and synthesizing new electrolyte components (solvents, salts, and additives), but also in making use of already-existing components in an innovative manner. These can be

condensed into a few general approaches that have been effectively applied to a wide spectrum of battery chemistries from Si and Li^0 to multivalent metals and from improved intercalation to conversion-reaction materials.

Superconcentrating

Traditionally, electrolytes were confined to a salt concentration around 1 molarity (or 1 molality) to maximize ion transport. This limit was lifted after researchers realized that an extra-high salt concentration could bring a series of benefits that offset the loss in ion transport due to the viscosity increase (*52*, *53*). These benefits include nonflammability, anticorrosion at high voltage, high cation transference, and, most importantly, new interphasial chemistries. The rationale underneath such changes lies in the solvation sheath structures of both cation and anion, which are significantly altered at these unusual salt concentrations. One extreme example is the WiSE (*23*–*25*), in which 21 molality or higher lithium salts dissolved in water enable the formation of an SEI in water and expand the electrochemical stability window of aqueous electrolytes from 1.23 V into the regime of >3.0 V to support high-voltage aqueous LIBs. At such extra-high salt concentrations, not only are solvation sheaths restructured, but in extended length scales, the hydrogen bonding between water molecules completely disappears and an anion-cation network forms. The superconcentration concept applies to aqueous, nonaqueous, and even hybrid electrolyte systems. One particular modification is the so-called localized high-concentration electrolytes, in which a noncoordinating co-solvent (usually polyfluorinated ethers) is introduced to further disrupt the liquid structure of the parent electrolyte so that the overall salt concentration would be near 1.0 M but locally the cation still sees a solvation structure similar to a superconcentration regime (*54*).

Singularizing

Because it has been widely established that interphases formed in situ on an electrode are highly inhomogeneous in both chemistry and morphology, efforts are being made to identify which chemical is the single most effective ingredient in ensuring cell reversibility. On the basis of this knowledge, efforts could then be made to enrich such chemistry in interphases or even make the interphases in singular composition. Currently, the attention has been focused on two distinct inorganic salts, LiF and Li_2O, as effective yet manipulable SEI components.

The presence of various fluorides, either organic or inorganic, has been long identified in both SEIs and CEIs, but no simple linear relation could be established between fluoride content and interphase performance. Fluorine is the most resistive element toward both oxidation and reduction, so it should make an ideal ingredient for interphases designed to support aggressive battery chemistries. However, fluorides are excellent insulators not only to electrons but also to ions, so it is counterintuitive that a pure fluoride-based interphase could function at all. Simulations reveal that when fluorides interface with other impurities such as oxides or carbonates at nanometric scale, highly conductive pathways for Li^+ can be created along the grain boundaries (*55*). It becomes increasingly apparent that fluoride distribution in the interphases and interactions with other species are more critical than the fluoride content alone (*56*). Thus far, knowledge about precisely tailoring the nanometric morphology of these fluoride species is still missing. A semi-empirical approach adopted as the design philosophy is that fluorine must be prestored in the structures of various solvents, additives, or anions, and would then become available only when electrochemically activated. Fluorides formed in this manner would exist in nanometric scales and well interfaced with other interphasial ingredients. Highly fluorinated interphases have been made with impressive performances, examples of which include the almost pure LiF- or NaF-based SEIs formed in WiSE from TFSI anions and the LiF-enriched CEIs formed in fluorinated ether electrolytes (*23*, *24*).

Li_2O was identified recently as an alternative interphasial ingredient on Li^0 on the basis of observation and analysis achieved under cryogenic electron microscopy (*57*, *58*). Unexpectedly, highly fluorinated ether electrolytes that have led to fluorinated CEI on a NMC cathode simultaneously generated Li_2O-based, monolithic SEIs on a Li^0 surface without any detectable fluorides (*58*, *59*). More pure interphases of singular chemical composition were thus attempted, in which both LiF and Li_2O were artificially applied on Li^0 surfaces using a suspension of corresponding nanoparticles (*60*). It is anticipated that with increasing understanding about interphasial chemistry and morphology, we will witness further such efforts with new chemical ingredients. However, caution should be exercised here because we actually know very little about how LiF or Li_2O functions in an interphase. The correlation between their presence in interphases and their effect on electrochemical performance remains superficial and highly empirical. More fundamental knowledge is needed before we can determine whether an ideal interphase should be singular or heterogeneous.

Solidifying

Making electrolytes solid has always been desirable because solid electrolytes (SEs) provide nonflammability, zero leakage, and mechanical resistance against Li^0-dendrite growth, thus offering more convincing feasibility of enabling Li^0 anodes in wide range of operating temperatures (*61*). Decades ago, researchers started seeking suitable materials with an excellent electrolyte nature, i.e., high ionic conductivity (σ_{Li+} > 0.1 mS/cm) but low electronic conductivity ($\sigma_e < 10^{-7}$ mS/cm). These efforts have covered organic materials such as polymeric, inorganic ceramic or glassy materials such as oxides, halides, and sulfides as well as composites that combine all of them (*61*, *62*).

Polymeric ion conductors are of interest because of their flexibility and processibility, but the coupling between the segmental motion of the oligoether chain and the ion transport constitutes an insurmountable barrier to achieving sufficient ion conductivities at ambient temperature (*63*). The average conductivity of ~10^{-5} S/cm at 25°C restricts the use of these polymer electrolytes in batteries to niche applications at elevated temperatures, such as the solid lithium-metal batteries adopted in certain electric vehicles in Europe. Fundamental breakthroughs are needed to decouple ion transport from polymeric segmental relaxation, which requires a new ion transport mechanism. One recent advance in which cellulose chains are artificially expanded by bridging copper ions so that Li^+ can travel independently from the polymeric chain represents one such direction (*64*).

Among the inorganic solid materials, sulfides have emerged as a top contender, with high ionic conductivities at room temperature (>10^{-3} S/cm) and high compliance to be processed into ultrathin dense layers, but their moderate electrochemical stability requires extra interphasial protection against extreme potentials of numerous cathode and anode chemistries (*61*). Argyrodite Li_6PS_5Cl (LPSCl) has been revealed to form interphases with both Li^0 and Si, which is dominated by electronic insulating yet ionic conductive salts such as LiCl, Li_2S, and Li_3P (*65*). For high-voltage cathodes such as layered transition-metal oxides, sulfides are usually protected by artificial interphases that are formed by coating on sulfide SEs a layer of ceramic with high oxidative stability, examples of which are lithium niobates and lithium borates. With tailored composition and interfacial engineering, sulfide-based SEs could work at voltages >4.2 V (*66*).

Various types of scalable, low-cost, uniform coating strategies on cathode materials are still being explored, because the mismatch in chemical and mechanical properties between the cathode active materials and the SE could lead to uncontrolled impedance growth that negatively affects cell performance (*67*). Coating layers not only chemically and electrochemically protects the interfaces from decomposition, but also serves as a mechanical buffer to stabilize solid-solid interfaces. The formation of dendritic

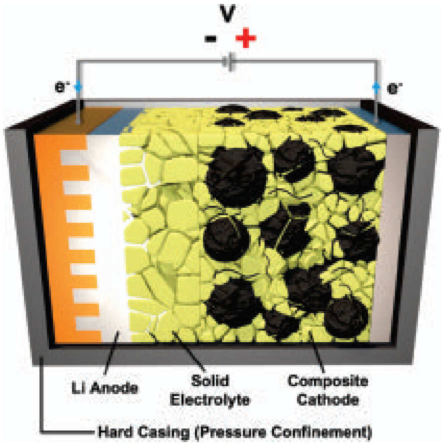


Fig. 4. Interfacing solid with solid. Interfacing solid materials in an all-solid battery encounters both challenges and opportunities with the solid-solid contact points of active anode/cathode materials and the SEs.

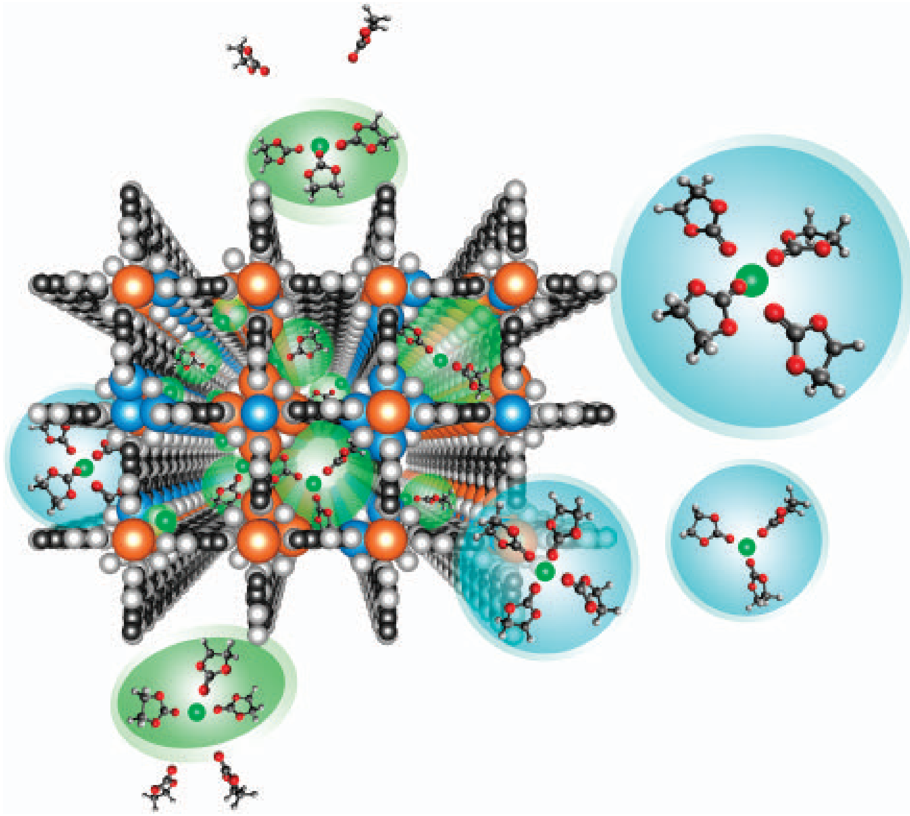


Fig. 5. Solvation in nanoconfinement. When the host environment becomes comparable in size with the ion solvation sheaths, forced interaction between the host and ion or its solvent members becomes inevitable, which induces partial or complete desolvation of the ions and generates a series of unusual properties.

and dead Li^0 continues to plague Li metal anodes because of imperfections and inhomogeneity at metal-SE interface, leading to these dangerous Li^0 morphologies that become exacerbated at higher current densities. Recently, the inhomogeneous stripping of lithium and the subsequent formation of voids have also become a focus of research (*68*). Plastic deformation of the metal through mechanical compression shows promise in preventing dendritic and dead Li^0 (*69*, *70*). However, obtaining these pressures may require significant cell-level penalties because of the need for special fixturing.

The paramount challenge for SEs still comes from solid-solid interfaces, because SEs cannot flow or infiltrate into the porosity structure of the electrodes. A common approach to overcoming the physical contact issue is the addition of a controlled amount of liquid electrolytes to fill the voids and gaps. This renders the system less solid and results in a hybrid electrolyte or semi-SE (*71*). Conversely, the solid-solid contact challenge could also be turned into an advantage to enable three-dimensional (3D) architectures for the Li or Si anodes, because the interphasial reactions can only occur at those contact points, whereas large volume expansion and shrinkage can be accommodated with appropriate 3D design and dynamic pressure control (Fig. 4) (*72*). The single-ion conducting nature of SEs eliminates concentration polarization and allows accommodation of local effects that cause nonuniform reaction and eventual capacity fade (*73*). Although it remains a key hurdle for the battery field to quantify the coupled electrochemical-mechanical phenomena, recent work has been rather encouraging, showing a stable and long cycle-life of laboratory-scale batteries.

Liquefying

In addition to liquids and solids, a new frontier has been created in which a quasi-gases are harnessed as electrolyte solvents. These efforts enable operation at lower temperatures not otherwise possible with typical electrolytes. Relatively polar gaseous hydrofluorocarbon molecules such as fluoromethane (CH_3F) can be liquefied under moderate pressure (*74*), which dissolves salts at room temperature to form liquefied gas electrolytes. These electrolytes are nontoxic and have relatively strong covalent bonds that allow for a wide electrochemical window. Because of their small molecular size and weak intermolecular attraction, these solvents display exceptionally low viscosity that enables a superior solvent dielectric-fluidity factor that is higher than that of typical electrolyte solvents such as esters and ethers. The solvents also have extremely low melting points that render batteries operational below –60°C. Additives such as tetrahydrofuran (THF) (*75*) can be introduced to alter the Li^+-solvation structure, which has been fully populated with CH_3F molecules. The intrinsic fluorine-rich chemistry and the self-equilibrated pressure enable Li^0 reversibility even at very high current densities (10 mA/cm^2) and low temperatures. Other types of co-solvent could also be used to widen the operation temperature range, especially at the upper limit, where salt precipitates and ion transport shutdown occurs. More recent use of difluoromethane (CH_2F_2) in combination with dimethoxyethane co-solvent have shown such an improvement (*76*), and the concept was also extended to multivalent chemistries such as Zn (*77*). These unconventional electrolytes, although confined to limited applications because they require a hermetical stainless steel case, have significant potential for further development.

Nano-confining

The size of most ion solvation sheaths ranges from an angstrom to nearly a nanometer, depending on the solvent molecules and the solvation numbers. With the emergence of diversified nanomaterials, an interesting question arises: How would the solvated ions behave in these sub-nano structures? There have been numerous reports on the unusual properties of electrolytes in nano-confinement. For example, anomalous capacitance arose when solvated tetraalkyl ammonium cations were forced into pores as small as 0.65 nm (*78*). Significantly different interphasial behaviors and morphologies of Li^0 anode emerged when infusing conventional LIB electrolytes

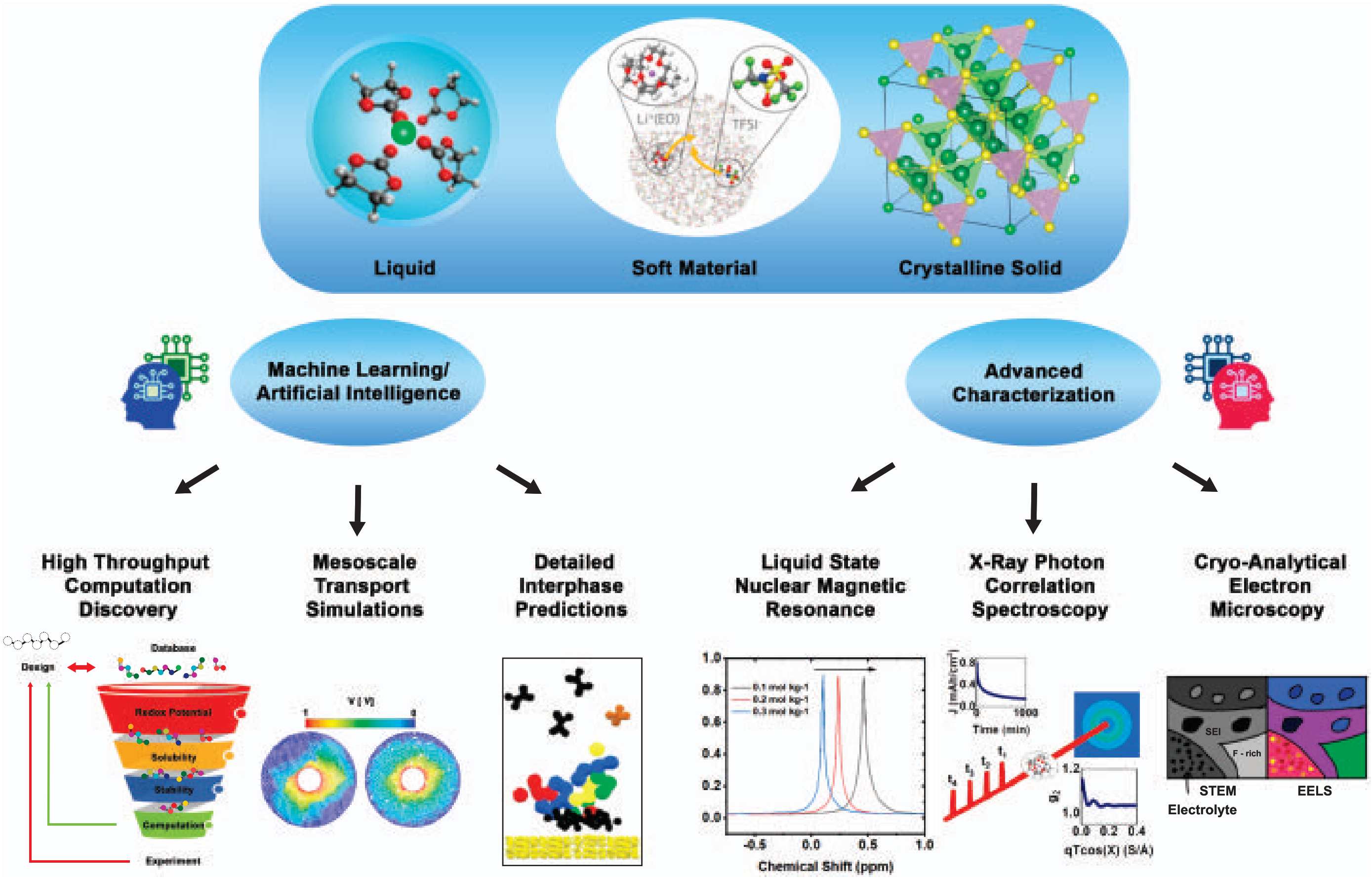


Fig. 6. Seeing electrolyte dynamics at molecular resolution. Advanced ex situ and in situ/in operando characterizations assisted by computational simulation and modeling bring a new horizon to the understanding and discovery of electrolytes and their interphases.

into nanopores of a ceramic-polymer composite host of mean diameter ~40 nm (*79*). The electrochemical stability window was significantly expanded when ether-based electrolytes were frozen into the sub-nano channels of 0.29 nm created by metal-organic framework structures (*80*, *81*). Finally, extraordinarily fast ion transport far above the corresponding bulk polymer electrolyte occured along the ceramic-polymer interfaces when the same polymer was embedded into the channels of a ceramic host as small as 40 nm in diameter (*82*). Although these scattered phenomena do not seem to be related at first glimpse, and the investigators in each study offered widely different rationales, there is a converging factor in all of them: In the nanostructured environment, the solvated ions were likely to interact with the surfaces of the host materials (Fig. 5), during which partial or even complete desolvation of the ions occured, producing electrolytes at chemical states that are not fully understood yet. Therefore, there might be opportunities to leverage these conditions.

Outlook: A bright future with apprehensions

A battery is a multicomponent system consisting of at minimum a positive electrode (cathode), a negative electrode (anode), and an electrolyte. For the battery chemistry to work, all of these components must synchronize with each other. The failure of a battery often occurs not as a result of a single malfunctioning component but rather because of the interactions between different components.

Constrained by such a synchronization requirement, the electrolyte has often been the languishing component (*8*) because it must simultaneously interface with every other component in a battery, be it active components such as the cathode and anode or inactive components such as separators, current collectors, and cell packaging. A significant amount of knowledge has been generated about electrolytes and their relation to interphase. Nevertheless, many basic questions remain unanswered, such as: (i) How exactly do ions travel across interphases consisting of intrinsically insulating materials? (ii) How can one precisely measure electronic and ion conductivity within the interphases? (iii) Which components are effective and functional in the interphases and which are unnecessary and redundant? This list of questions will surely grow as new systems are developed and explored. In the meantime, new insights are being provided by various analytical methods that were previously unavailable. For example, combined chemical analysis, cryogenic electron microscopy, and pair distribution function (PDF) acquired from a synchrotron revealed that the interphasial chemistry on Li^0 is much more complicated than was once believed, with the exotic interphasial components such as glassy Li^0 and lithium hydride still under debate (*83*, *84*).

To answer all of these questions, characterization techniques, especially those operational under in situ/in operando conditions and those with extremely high spatial and temporal resolutions, are needed. In the past decades, although many advanced characterization tools have been developed to probe and quantify the interphasial species, tools that directly probe the electrolyte are still absent. Given the noncrystalline and fast ion dynamics nature of typical electrolytes, spectroscopic methods with high chemical sensitivity and total scattering for local structure are most suitable because they are tolerant to long-range disordered arrangements. As shown in Fig. 6, a comprehensive understanding about solvation structure in electrolytes, the associated interphase chemistry, and its distribution along 2D and 3D length scales, as well as how the chemistry evolves

during electrochemical processes, could be achieved by the study of large datasets collected using a combination of a host of techniques. These include liquid and solid nuclear magnetic resonance (*13*), differential PDF derived from x-ray or neutron scattering (*85*), and cryogenic analytical electron microscopy equipped with tomography and electron energy loss spectroscopy (*86*).

As highly powerful supercomputing capabilities become more accessible, complicated processes in electrolyte bulk and interfaces are increasingly available for modeling with improved accuracy. Computational methods increasingly play a role in the discovery of new electrolytes and additives by using density functional theory to screen the thermodynamic oxidative and reductive stability of various organic molecules (*87*), in predicting ion mobility in SEs using methods such as nudged elastic band coupled with ab initio molecular dynamics (*88*), as well as in understanding interfacial and interphasial processes (*89*, *90*). These approaches have led to the development of infrastructures and associated databases to provide electrolyte information to the wider research community (*91*). With the increased interest in automated high-throughput experimentation (so-called robotic labs), these computation screening approaches are now aided by large volumes of experimental data, lending themselves to analysis using machine learning approaches (*92*).

Computation predictions complement experimental characterization methods by offering insight into the detailed and transient processes at interfaces. The recent molecular dynamic simulations of WiSE and its derivatives on electrified surfaces have revealed the possibility of predicting interphasial chemistries based on how the solvents and ions assemble and position in the inner-Helmholtz layers of the electrode before interphase formation (*89*). Further, molecular dynamic simulations also shed light on the reduction reactions responsible for interphase formation producing solid and gaseous products including C_2H_4, Li_2CO_3, LiF, and LEDC (*93*).

Finally, the increasing interest in developing electrolytes with enhanced transport properties has shed light on scale-bridging approaches linking ion correlations at the atomic scale to transport properties at the continuum scale using Stefan-Maxwell formalism (*94*). These methods have recently incorporated additional phenomena such as solvent motion (*95*) and its role in the precise definition of ion transference and the role of morphological effects at the nanoscale on conductivity (*96*), both of which use large-scale computations.

Although computational methods have blossomed over the past decade, further efforts are needed to move beyond thermodynamic calculations into the prediction of chemical reactions so that the cascade of reactions leading to the formation of interphases can be precisely described and tighter coupling between experimental methods and predictions can be forged. Artificial intelligence and machine learning methods will eventually grow to a leading role in these endeavors. Validation of modeling requires further support from advanced characterization.

The past decade has witnessed the explosive growth of research activities on battery chemistry and materials, with electrolyte development taking a substantial role. This emphasis on an area of vital importance to the future of our planet and civilization will be sustained and driven further by anxieties over both geopolitical and supply chain risks. Given that a battery is a system consisting of multiple components, complications often arise when evaluating a single component in different environments and settings. To enforce the reproducibility and comparability of the battery performance data in the literature, rigorous practice and standardized protocols by researchers are strongly encouraged. Some efforts have already been made in this direction (*97*, *98*), and we view these as indispensable guidelines that the battery, chemistry, and materials research community should follow strictly.

REFERENCES AND NOTES

1. J. O.'M. Bockris, A. K. N. Reddy, *Modern Electrochemistry 1: Ionics*, ed. 2. (Plenum Press, 1998).
2. M. Winter, B. Barnett, K. Xu, Before Li ion batteries. *Chem. Rev.* **118**, 11433–11456 (2018). doi: 10.1021/acs.chemrev.8b00422; pmid: 30500179
3. K. Xu, Nonaqueous liquid electrolytes for lithium-based rechargeable batteries. *Chem. Rev.* **104**, 4303–4418 (2004). doi: 10.1021/cr030203g; pmid: 15669157
4. E. A. Peled, The electrochemical behavior of alkali and alkaline earth metals in nonaqueous battery systems: The solid electrolyte interphase model. *J. Electrochem. Soc.* **126**, 2047–2051 (1979). doi: 10.1149/1.2128859
5. D. R. Gallus, R. Wagner, S. Wiemers-Meyer, M. Winter, I. Cekic-Laskovic, New insights into the structure-property relationship of high-voltage electrolyte components for lithium-ion batteries using the pK_a value. *Electrochim. Acta* **184**, 410–416 (2015). doi: 10.1016/j.electacta.2015.10.002
6. M. Winter, The solid electrolyte interphase–the most important and the least understood solid electrolyte in rechargeable Li batteries. *Z. Phys. Chem.* **223**, 1395–1406 (2009). doi: 10.1524/zpch.2009.6086
7. J. B. Goodenough, Y. Kim, Challenges for rechargeable Li batteries. *Chem. Mater.* **22**, 587–603 (2010). doi: 10.1021/cm901452z
8. K. Xu, Li ion battery electrolytes. *Nat. Energy* **6**, 763 (2021). doi: 10.1038/s41560-021-00841-6
9. K. Xu, Y. Lam, S. S. Zhang, T. R. Jow, T. B. Curtis, Solvation sheath of Li^+ in nonaqueous electrolytes and its implication of graphite/electrolyte interface chemistry. *J. Phys. Chem. C* **111**, 7411–7421 (2007). doi: 10.1021/jp068691u
10. M. Onuki *et al*., Identification of the source of evolved gas in Li-ion batteries using ^{13}C-labeled solvents. *J. Electrochem. Soc.* **155**, A794–A797 (2008). doi: 10.1149/1.2969947
11. G. V. Zhuang, K. Xu, H. Yang, T. R. Jow, P. N. RossJr.., Lithium ethylene dicarbonate identified as the primary product of chemical and electrochemical reduction of EC in 1.2 M $LiPF_6$/EC:EMC electrolyte. *J. Phys. Chem. B* **109**, 17567–17573 (2005). doi: 10.1021/jp052474w; pmid: 16853247
12. A. von Cresce, K. Xu, Preferential solvation of Li^+ directs formation of interphase on graphitic anode. *Electrochem. Solid-State Lett.* **14**, A154–A156 (2011). doi: 10.1149/1.3615828
13. X. Bogle, R. Vazquez, S. Greenbaum, A. W. Cresce, K. Xu, Understanding Li^+-solvent interaction in nonaqueous carbonate electrolytes with ^{17}O NMR. *J. Phys. Chem. Lett.* **4**, 1664–1668 (2013). doi: 10.1021/jz400661k; pmid: 26282976
14. T. Liu *et al*., In situ quantification of interphasial chemistry in Li-ion battery. *Nat. Nanotechnol.* **14**, 50–56 (2019). doi: 10.1038/s41565-018-0284-y; pmid: 30420761
15. L. Wang *et al*., Identifying the components of the solid-electrolyte interphase in Li-ion batteries. *Nat. Chem.* **11**, 789–796 (2019). doi: 10.1038/s41557-019-0304-z; pmid: 31427766
16. L. Ma *et al*., A guide to ethylene carbonate-free electrolyte making for Li-ion cells. *J. Electrochem. Soc.* **164**, A5008–A5018 (2017). doi: 10.1149/2.0191701jes
17. S. Klein *et al*., Understanding the outstanding high-voltage performance of NCM523||graphite lithium ion cells after elimination of ethylene carbonate solvent from conventional electrolyte. *Adv. Energy Mater.* **11**, 2003738 (2021). doi: 10.1002/aenm.202003738
18. A. Würsig, H. Buqa, M. Holzapfel, F. Krueich, P. Novák, Film formation at positive electrodes in lithium-ion batteries. *Electrochem. Solid-State Lett.* **8**, A34–A37 (2005). doi: 10.1149/1.1836114
19. B. Streipert *et al*., Conventional electrolyte and inactive electrode materials in lithium-ion batteries: Determining cumulative impact of oxidative decomposition at high voltage. *ChemSusChem* **13**, 5301–5307 (2020). doi: 10.1002/cssc.202001530; pmid: 32692891
20. N. Dupré *et al*., Aging of the $LiNi_{1/2}Mn_{1/2}O_{1/2}$ positive electrode interface in electrolyte. *J. Electrochem. Soc.* **156**, C180–C185 (2009). doi: 10.1149/1.3098494
21. A. von Cresce, K. Xu, Electrolyte additive in support of 5 V Li ion chemistry. *J. Electrochem. Soc.* **158**, A337–A342 (2011). doi: 10.1149/1.3532047
22. L. Trahey *et al*., Energy storage emerging: A perspective from the Joint Center for Energy Storage Research. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 12550–12557 (2020). doi: 10.1073/pnas.1821672117; pmid: 32513683
23. L. Suo *et al*., "Water-in-salt" electrolyte enables high-voltage aqueous lithium-ion chemistries. *Science* **350**, 938–943 (2015). doi: 10.1126/science.aab1595; pmid: 26586759
24. L. Suo *et al*., Advanced high-voltage aqueous lithium-ion battery enabled by "water-in-bisalt" electrolyte. *Angew. Chem. Int. Ed.* **55**, 7136–7141 (2016). doi: 10.1002/anie.201602397; pmid: 27120336
25. Y. Yamada *et al*., Hydrate-melt electrolytes for high-energy-density aqueous batteries. *Nat. Energy* **1**, 16129 (2016). doi: 10.1038/nenergy.2016.129
26. N. Yabuuchi, K. Kubota, M. Dahbi, S. Komaba, Research development on sodium-ion batteries. *Chem. Rev.* **114**, 11636–11682 (2014). doi: 10.1021/cr500192f; pmid: 25390643
27. T. Hosaka, K. Kubota, A. S. Hameed, S. Komaba, Research development on K-ion batteries. *Chem. Rev.* **120**, 6358–6466 (2020). doi: 10.1021/acs.chemrev.9b00463; pmid: 31939297
28. J. Muldoon *et al*., Electrolyte roadblocks to a magnesium rechargeable battery. *Energy Environ. Sci.* **5**, 5941–5950 (2012). doi: 10.1039/c2ee03029b
29. M. E. Arroyo-de Dompablo, A. Ponrouch, P. Johansson, M. R. Palacín, Achievements, challenges, and prospects of calcium batteries. *Chem. Rev.* **120**, 6331–6357 (2020). doi: 10.1021/acs.chemrev.9b00339; pmid: 31661250
30. M.-C. Lin *et al*., An ultrafast rechargeable aluminium-ion battery. *Nature* **520**, 325–328 (2015). doi: 10.1038/nature14340; pmid: 25849777
31. U. Kasavajjula, C. Wang, A. J. Appleby, Nano- and bulk-silicon-based insertion anodes for lithium-ion secondary cells. *J. Power Sources* **163**, 1003–1039 (2007). doi: 10.1016/j.jpowsour.2006.09.084
32. C. K. Chan *et al*., High-performance lithium battery anodes using silicon nanowires. *Nat. Nanotechnol.* **3**, 31–35 (2008). doi: 10.1038/nnano.2007.411; pmid: 18654447
33. M. N. Obrovac, V. L. Chevrier, Alloy negative electrodes for Li-ion batteries. *Chem. Rev.* **114**, 11444–11502 (2014). doi: 10.1021/cr500207g; pmid: 25399614
34. W. Xu *et al*., Lithium metal anodes for rechargeable batteries. *Energy Environ. Sci.* **7**, 513–537 (2014). doi: 10.1039/C3EE40795K
35. C. Fang *et al*., Quantifying inactive lithium in lithium metal batteries. *Nature* **572**, 511–515 (2019). doi: 10.1038/s41586-019-1481-z; pmid: 31435056
36. A. J. Louli *et al*., Diagnosing and correcting anode-free cell failure via electrolyte and morphological analysis. *Nat. Energy* **5**, 693–702 (2020). doi: 10.1038/s41560-020-0668-8
37. Y. Liang, H. Dong, D. Aurbach, Y. Yao, Current status and future directions of multivalent metal-ion batteries. *Nat. Energy* **5**, 646–656 (2020). doi: 10.1038/s41560-020-0655-0

38. G. R. Pastel *et al*., A sobering examination of the feasibility of aqueous aluminum batteries. *Energy Environ. Sci.* **15**, 2460–2469 (2022). doi: 10.1039/D2EE00134A
39. Z. Yu *et al*., Solvation structure and dynamics of $Mg(TFSI)_2$ aqueous electrolyte. *Energy Environ. Mater.* **5**, 295–304 (2021). doi: 10.1002/eem2.12174
40. S. B. Son *et al*., An artificial interphase enables reversible magnesium chemistry in carbonate electrolytes. *Nat. Chem.* **10**, 532–539 (2018). doi: 10.1038/s41557-018-0019-6; pmid: 29610460
41. L. Ma *et al*., Realizing high zinc reversibility in rechargeable batteries. *Nat. Energy* **5**, 743–749 (2020). doi: 10.1038/s41560-020-0674-x
42. P. Albertus, S. Babinec, S. Litzelman, A. Newman, Status and challenges in enabling the lithium metal electrode for high-energy and low-cost rechargeable batteries. *Nat. Energy* **3**, 16–21 (2018). doi: 10.1038/s41560-017-0047-2
43. W. Sun *et al*., A rechargeable zinc-air battery based on zinc peroxide chemistry. *Science* **371**, 46–51 (2021). doi: 10.1126/science.abb9554; pmid: 33384369
44. L. Yin *et al*., High performance printed AgO-Zn rechargeable battery for flexible electronics. *Joule* **5**, 228–248 (2021). doi: 10.1016/j.joule.2020.11.008
45. Y. K. Sun *et al*., Nanostructured high-energy cathode materials for advanced lithium batteries. *Nat. Mater.* **11**, 942–947 (2012). doi: 10.1038/nmat3435; pmid: 23042415
46. N.-S. Choi, J.-G. Han, S.-Y. Ha, I. Pak, C.-K. Back, Recent advances in the electrolytes for interfacial stability of high-voltage cathodes in lithium-ion batteries. *RSC Advances* **5**, 2732–2748 (2015). doi: 10.1039/C4RA11575A
47. D. H. Seo *et al*., The structural and chemical origin of the oxygen redox activity in layered and cation-disordered Li-excess cathode materials. *Nat. Chem.* **8**, 692–697 (2016). doi: 10.1038/nchem.2524; pmid: 27325096
48. F. Wu, G. Yushin, Conversion cathodes for rechargeable lithium and lithium-ion batteries. *Energy Environ. Sci.* **10**, 435–459 (2017). doi: 10.1039/C6EE02326F
49. X. Fan *et al*., High energy-density and reversibility of iron fluoride cathode enabled via an intercalation-extrusion reaction. *Nat. Commun.* **9**, 2324 (2018). doi: 10.1038/s41467-018-04476-2; pmid: 29899467
50. A. Manthiram, Y. Fu, S.-H. Chung, C. Zu, Y.-S. Su, Rechargeable lithium-sulfur batteries. *Chem. Rev.* **114**, 11751–11787 (2014). doi: 10.1021/cr500062v; pmid: 25026475
51. T. Liu *et al*., Current challenges and routes forward for nonaqueous lithium–air batteries. *Chem. Rev.* **120**, 6558–6625 (2020). doi: 10.1021/acs.chemrev.9b00545; pmid: 32090540
52. M. Li, C. Wang, Z. Chen, K. Xu, J. Lu, New concepts in electrolytes. *Chem. Rev.* **120**, 6783–6819 (2020). doi: 10.1021/acs.chemrev.9b00531; pmid: 32022546
53. O. Borodin, J. Self, K. A. Persson, C. Wang, K. Xu, Uncharted waters: Super-concentrated electrolytes. *Joule* **4**, 69–100 (2020). doi: 10.1016/j.joule.2019.12.007
54. S. Chen *et al*., High-voltage lithium-metal batteries enabled by localized high-concentration electrolytes. *Adv. Mater.* **30**, e1706102 (2018). doi: 10.1002/adma.201706102; pmid: 29575163
55. Q. Zhang *et al*., Synergetic effects of inorganic components in solid electrolyte interphase on high cycle efficiency of lithium ion batteries. *Nano Lett.* **16**, 2011–2016 (2016). doi: 10.1021/acs.nanolett.5b05283; pmid: 26889564
56. C. Wang, Y. S. Meng, K. Xu, Fluorinating interphases. *J. Electrochem. Soc.* **166**, A5184–A5186 (2018). doi: 10.1149/2.0281903jes
57. X. Wang *et al*., New insights on the structure of electrochemically deposited lithium metal and its solid electrolyte interphases via cryogenic TEM. *Nano Lett.* **17**, 7606–7612 (2017). doi: 10.1021/acs.nanolett.7b03606; pmid: 29090936
58. Y. Li *et al*., Atomic structure of sensitive battery materials and interfaces revealed by cryo-electron microscopy. *Science* **358**, 506–510 (2017). doi: 10.1126/science.aam6014; pmid: 29074771
59. R. Guo, B. Gallant, Li_2O solid electrolyte interphase: Probing transport properties at the chemical potential of lithium. *Chem. Mater.* **32**, 5525–5533 (2020). doi: 10.1021/acs.chemmater.0c00333
60. M. S. Kim *et al*., Suspension electrolyte with modified Li^+ solvation environment for lithium metal batteries. *Nat. Mater.* **21**, 445–454 (2022). doi: 10.1038/s41563-021-01172-3; pmid: 35039645
61. A. Manthiram, X. Yu, S. Wang, Lithium battery chemistries enabled by solid-state electrolytes. *Nat. Rev. Mater.* **2**, 16103 (2017). doi: 10.1038/natrevmats.2016.103
62. Z. Zou *et al*., Mobile ions in composite solids. *Chem. Rev.* **120**, 4169–4221 (2020). doi: 10.1021/acs.chemrev.9b00760; pmid: 32267697
63. K. Xu, Electrolytes and interphases in Li-ion batteries and beyond. *Chem. Rev.* **114**, 11503–11618 (2014). doi: 10.1021/cr500003w; pmid: 25351820
64. C. Yang *et al*., Copper-coordinated cellulose ion conductors for solid-state batteries. *Nature* **598**, 590–596 (2021). doi: 10.1038/s41586-021-03885-6; pmid: 34671167
65. Y. Zhu, X. He, Y. Mo, First principles study on electrochemical and chemical stability of solid electrolyte–electrode interfaces in all-solid-state Li-ion batteries. *J. Mater. Chem. A Mater. Energy Sustain.* **4**, 3253–3266 (2016). doi: 10.1039/C5TA08574H
66. A. Banerjee, X. Wang, C. Fang, E. A. Wu, Y. S. Meng, Interfaces and interphases in all-solid-state batteries with inorganic solid electrolytes. *Chem. Rev.* **120**, 6878–6933 (2020). doi: 10.1021/acs.chemrev.0c00101; pmid: 32603100
67. Y. Xiao, L. J. Miara, Y. Wang, G. Ceder, Computational screening of cathode coatings for solid-state batteries. *Joule* **3**, 1252–1275 (2019). doi: 10.1016/j.joule.2019.02.006
68. T. Krauskopf, H. Hartmann, W. G. Zeier, J. Janek, Toward a fundamental understanding of the lithium metal anode in solid-state batteries: An electrochemo-mechanical study on the garnet-type solid electrolyte $Li_{6.25}Al_{0.25}La_3Zr_2O_{12}$. *ACS Appl. Mater. Interfaces* **11**, 14463–14477 (2019). doi: 10.1021/acsami.9b02537; pmid: 30892861
69. C. Fang *et al*., Pressure-tailored lithium deposition and dissolution in lithium metal batteries. *Nat. Energy* **6**, 987–994 (2021). doi: 10.1038/s41560-021-00917-3
70. Y. Wang, T. Liu, J. Kumar, Effect of pressure on lithium metal deposition and stripping against sulfide-based solid electrolytes. *ACS Appl. Mater. Interfaces* **12**, 34771–34776 (2020). doi: 10.1021/acsami.0c06201; pmid: 32657569
71. L. R. Mangani, C. Villevieille, Mechanical *vs.* chemical stability of sulphide-based solid-state batteries. Which one is the biggest challenge to tackle? Overview of solid-state batteries and hybrid solid state batteries. *J. Mater. Chem. A Mater. Energy Sustain.* **8**, 10150–10167 (2020). doi: 10.1039/D0TA02984J
72. D. H. S. Tan *et al*., Carbon-free high-loading silicon anodes enabled by sulfide solid electrolytes. *Science* **373**, 1494–1499 (2021). doi: 10.1126/science.abg7217; pmid: 34554780
73. J. Gao, C. Wang, D.-W. Han, D.-M. Shin, Single-ion conducting polymer electrolytes as a key jigsaw piece for next-generation battery applications. *Chem. Sci.* **12**, 13248–13272 (2021). doi: 10.1039/D1SC04023E; pmid: 34777744
74. C. S. Rustomji *et al*., Liquefied gas electrolytes for electrochemical energy storage devices. *Science* **356**, eaal4263 (2017). doi: 10.1126/science.aal4263; pmid: 28619715
75. Y. Yang *et al*., High-efficiency lithium-metal anode enabled by liquefied gas electrolytes. *Joule* **3**, 1986 (2019). doi: 10.1016/j.joule.2019.06.008
76. D. M. Davies *et al*., A safer, wide-temperature liquefied gas electrolyte based on difluoromethane. *J. Power Sources* **493**, 229668 (2021). doi: 10.1016/j.jpowsour.2021.229668
77. L. Ma *et al*., High-efficiency zinc-metal anode enabled by liquefied gas electrolytes. *ACS Energy Lett.* **6**, 4426–4430 (2021). doi: 10.1021/acsenergylett.1c02084
78. J. Chmiola *et al*., Anomalous increase in carbon capacitance at pore sizes less than 1 nanometer. *Science* **313**, 1760–1763 (2006). doi: 10.1126/science.1132195; pmid: 16917025
79. Y. Lu, Z. Tu, L. A. Archer, Stable lithium electrodeposition in liquid and nanoporous solid electrolytes. *Nat. Mater.* **13**, 961–969 (2014). doi: 10.1038/nmat4041; pmid: 25108613
80. Z. Chang *et al*., A liquid electrolyte with de-solvated lithium ions for lithium-metal battery. *Joule* **4**, 1776–1789 (2020). doi: 10.1016/j.joule.2020.06.011
81. Z. Chang *et al*., Beyond the concentrated electrolyte: Further depleting solvent molecules within a Li^+ solvation sheath to stabilize high-energy-density lithium metal batteries. *Energy Environ. Sci.* **13**, 4122–4131 (2020). doi: 10.1039/D0EE02769C
82. X. Zhang *et al*., Vertically aligned and continuous nanoscale ceramic–polymer interfaces in composite solid polymer electrolytes for enhanced ionic conductivity. *Nano Lett.* **18**, 3829–3838 (2018). doi: 10.1021/acs.nanolett.8b01111; pmid: 29727578
83. C. Fang *et al*., Quantifying inactive lithium in lithium metal batteries. *Nature* **572**, 511–515 (2019). doi: 10.1038/s41586-019-1481-z; pmid: 31435056
84. Z. Shadike *et al*., Identification of LiH and nanocrystalline LiF in the solid-electrolyte interphase of lithium metal anodes. *Nat. Nanotechnol.* **16**, 549–554 (2021). doi: 10.1038/s41565-020-00845-5; pmid: 33510453
85. S. H. Lapidus *et al*., Solvation structure and energetics of electrolytes for multivalent energy storage. *Phys. Chem. Chem. Phys.* **16**, 21941–21945 (2014). doi: 10.1039/C4CP03015J; pmid: 25208768
86. M. J. Zachman, Z. Tu, S. Choudhury, L. A. Archer, L. F. Kourkoutis, Cryo-STEM mapping of solid-liquid interfaces and dendrites in lithium-metal batteries. *Nature* **560**, 345–349 (2018). doi: 10.1038/s41586-018-0397-3; pmid: 30111789
87. L. Cheng *et al*., Accelerating electrolyte discovery for energy storage with high-throughput screening. *J. Phys. Chem. Lett.* **6**, 283–291 (2015). doi: 10.1021/jz502319n; pmid: 26263464
88. A. M. Nolan, Y. Zhu, X. He, Q. Bai, Y. Mo, Computation-accelerated design of materials and interfaces for all-solid-state lithium-ion batteries. *Joule* **2**, 2016–2046 (2018). doi: 10.1016/j.joule.2018.08.017
89. O. Borodin *et al*., Modeling insight into battery electrolyte electrochemical stability and interfacial structure. *Acc. Chem. Res.***50**, 2886–2894 (2017). doi: 10.1021/acs.accounts.7b00486; pmid: 29164857
90. S. Hou *et al*., Solvation sheath reorganization enables divalent metal batteries with fast interfacial charge transfer kinetics. *Science* **374**, 172–178 (2021). doi: 10.1126/science.abg3954; pmid: 34618574
91. X. Qu *et al*., The Electrolyte Genome Project: A big data approach in battery materials discovery. *Comput. Mater. Sci.* **103**, 56–67 (2015). doi: 10.1016/j.commatsci.2015.02.050
92. A. Dave *et al*., Autonomous discovery of battery electrolytes with robotic experimentation and machine learning. *Cell Rep. Phys. Sci.* **1**, 100264 (2020). doi: 10.1016/j.xcrp.2020.100264
93. L. Alzate-Vargas *et al*., Insight into SEI growth in Li-ion batteries using molecular dynamics and accelerated chemical reactions. *J. Phys. Chem. C* **125**, 18588–18596 (2021). doi: 10.1021/acs.jpcc.1c04149
94. D. R. Wheeler, J. Newman, Molecular dynamics simulations of multicomponent diffusion. 1. Equilibrium method. *J. Phys. Chem. B* **108**, 18353–18361 (2004). doi: 10.1021/jp047850b
95. H.-G. Steinrück *et al*., Concentration and velocity profiles in a polymeric lithium-ion battery electrolyte. *Energy Environ. Sci.* **13**, 4312–4321 (2020). doi: 10.1039/D0EE02193H
96. M. Chintapalli, K. Higa, X. C. Chen, V. Srinivasan, N. Balsara, Simulation of local ion transport in lamellar block copolymer electrolytes based on electron micrographs. *J. Polym. Sci., B, Polym. Phys.* **55**, 266–274 (2017). doi: 10.1002/polb.24268
97. Y.-K. Sun, An experimental checklist for reporting battery performances. *ACS Energy Lett.* **6**, 2187–2189 (2021). doi: 10.1021/acsenergylett.1c00870
98. A. K. Stephan, Standardized battery reporting guidelines. *Joule* **5**, 1–2 (2021). doi: 10.1016/j.joule.2020.12.026

ACKNOWLEDGMENTS

We thank D. Cheng, S. Parab, M. Shen, and M. Zhang for help with graphics. **Funding:** K.X. and V.S. were supported by the Joint Center for Energy Storage Research (JCESR), an energy hub funded by US Department of Energy, Office of Science, Basic Energy Sciences (BES). Y.S.M. was also supported by JCESR under award number DE-SC0002357. **Competing interests:** The authors declare no competing interests. **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

Submitted 4 April 2022; accepted 28 October 2022
10.1126/science.abq3750

RESEARCH ARTICLE SUMMARY

SYSTEMS BIOLOGY

Principles of gene regulation quantitatively connect DNA to RNA and proteins in bacteria

Rohan Balakrishnan†, Matteo Mori†, Igor Segota, Zhongge Zhang, Ruedi Aebersold, Christina Ludwig, Terence Hwa*

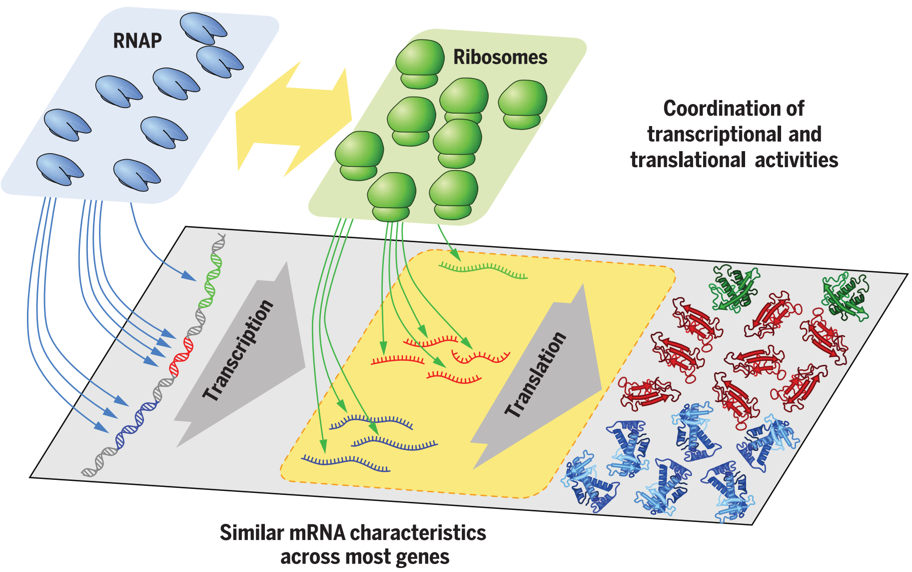
INTRODUCTION: The intracellular concentration of a protein depends on the rates of several processes, including transcription, translation, and the degradation and/or dilution of messenger RNAs (mRNAs) and proteins. These rates can be vastly different for different genes and across different growth conditions because of gene-specific regulation. At the systems level, protein concentrations are further affected by the availability of shared gene expression machineries—e.g., RNA polymerases and ribosomes—and are constrained by the approximately invariant cellular mass density. Even in one of the best-characterized model organisms, *Escherichia coli*, it is unclear how the gene-specific and systems-level effects work together toward setting the cellular proteome. This knowledge gap has not only hindered our efforts in building a predictive framework of gene expression but has also limited our abilities in guiding the rational design of gene circuits.

RATIONALE: We undertook a quantitative, genome-scale study, combining experimental and theoretical approaches, to tease apart the contribution of the specific and global effects on cellular protein concentrations in exponentially growing *E. coli* cells across a variety of growth conditions. We complemented genome-scale proteomic and transcriptomic data with biochemical measurements of total absolute mRNA abundances and synthesis rates. We compared these measurements to gene dosage and the concentrations of ribosomes and RNA polymerases to quantitatively characterize the activity of the gene expression machinery across conditions. This comprehensive dataset allowed us to analyze, in quantitative detail, the interplay between the activity of gene expression machinery, the activity of individual promoters, and the resulting protein concentrations.

RESULTS: We compiled a comprehensive atlas of the determinants of gene expression across conditions—from the concentrations of genes, mRNAs, and proteins to the rates of transcriptional and translational initiation and mRNA degradation for thousands of genes. We were able to determine the on rate of each promoter, a quantity capturing the overall effect of transcriptional regulation that has been elusive through most existing gene expression studies. Unexpectedly, we found that for most genes, the cytosolic protein concentrations were primarily determined by the innate magnitude of their promoter on rates, which spanned more than three orders of magnitude. Changes in protein concentrations resulting from changes in growth conditions were typically much smaller—well within one order of magnitude—and were mostly exerted through changes in transcription initiation.

E. coli's strategy to implement gene regulation can be summarized by two design principles. First, protein concentrations are predominantly set transcriptionally, with relatively invariant posttranscriptional characteristics (translation efficiencies and degradation rates) for most mRNAs and growth conditions. Second, the overall fluxes of transcription and translation are tightly coordinated: The average density of five ribosomes per kilobase is nearly invariant across mRNA species and across growth conditions, even though the mRNA and ribosome abundances can each vary substantially. We find this coordination to be implemented through the anti-sigma factor Rsd, which modulates the availability of RNA polymerases for transcription across different growth conditions. These two principles lead to a quantitative formulation of the central dogma of bacterial gene expression, connecting mRNA and protein concentrations to the regulatory activities of the corresponding promoters.

CONCLUSION: These quantitative relationships reveal the unexpectedly simple strategies used by *E. coli* to attain desired protein concentrations despite the complexity of global physiological constraints: Individual protein concentrations are primarily set by gene-specific transcriptional regulation, with global transcriptional regulation set to cancel the strong growth rate dependence of protein synthesis. These relations provide the basis for understanding the behavior of more complex genetic circuits in different conditions and for the inverse problem of deducing regulatory activities given the observed mRNA and protein levels. ■



Principles governing gene expression in *E. coli*. The availability of RNA polymerases (RNAPs) is coordinated with that of ribosomes, and the mRNA characteristics (translation efficiency and degradation rates) are uniform across most genes and growth conditions. These two principles prescribe a simple gene expression strategy in which protein concentrations are almost entirely controlled at the promoter level for most genes.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: hwa@ucsd.edu
†These authors contributed equally to this work.
Cite this article as R. Balakrishnan *et al.*, *Science* 378, eabk2066 (2022). DOI: 10.1126/science.abk2066

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abk2066

RESEARCH ARTICLE

SYSTEMS BIOLOGY

Principles of gene regulation quantitatively connect DNA to RNA and proteins in bacteria

Rohan Balakrishnan[1]†, Matteo Mori[1]†, Igor Segota[2], Zhongge Zhang[3], Ruedi Aebersold[4,5], Christina Ludwig[6], Terence Hwa[1,3]*

Protein concentrations are set by a complex interplay between gene-specific regulatory processes and systemic factors, including cell volume and shared gene expression machineries. Elucidating this interplay is crucial for discerning and designing gene regulatory systems. We quantitatively characterized gene-specific and systemic factors that affect transcription and translation genome-wide for *Escherichia coli* across many conditions. The results revealed two design principles that make regulation of gene expression insulated from concentrations of shared machineries: RNA polymerase activity is fine-tuned to match translational output, and translational characteristics are similar across most messenger RNAs (mRNAs). Consequently, in bacteria, protein concentration is set primarily at the promoter level. A simple mathematical formula relates promoter activities and protein concentrations across growth conditions, enabling quantitative inference of gene regulation from omics data.

Gene expression involves transcription of genes into mRNAs followed by translation of mRNAs into proteins. Protein concentrations are in turn determined by the balance between protein synthesis and dilution (Fig. 1A and fig. S1) for exponentially growing bacteria in which protein degradation is negligible (*1*, *2*). Intuitively, doubling the transcription initiation rate of a specific gene by an activator should result in doubling the concentration of the corresponding mRNA and protein in the absence of posttranscriptional regulation (*3*–*5*). However, if the transcription initiation rate of every gene were doubled, say by a global activator, protein concentrations could not possibly double: First, the synthesis of proteins is constrained by the translational capacity of the ribosomes (*6*, *7*). Second, in the well-characterized model bacterium *E. coli*, the total number of proteins per cell volume does not vary much across different growth conditions (fig. S2) (*8*, *9*), which makes it impossible to change the concentrations of most proteins in the same direction even if there is no constrain in gene expression capacity. Thus, the canonical single-gene picture of bacterial gene expression is not necessarily compatible with global constraints during broad changes in gene expression, as they occur, for example, upon changes in nutrient conditions or exposure to antibiotics. These effects make it difficult to link the transcriptional and translational regulation to the concentrations of mRNAs and proteins.

As the expression of each gene is ultimately determined by the rates at which the respective mRNAs and proteins are synthesized and diluted (fig. S1), we designed a battery of experiments to determine these rates by measuring the absolute mRNA and protein concentrations and their fluxes for *E. coli* growing exponentially under various conditions. Our findings establish characteristics of promoters and mRNAs that defied our expectations and reveal design principles underlying *E. coli*'s gene regulation program, which enable the cell to allocate its proteome in accordance to functional needs while complying with cellular constraints. We established a simple, quantitative relation that connects gene regulatory activities to mRNA and protein concentrations.

Results

Translation initiation rates are similar across mRNAs and growth conditions

A proteomics workflow was developed to quantify the abundance of individual *E. coli* proteins (*10*) by combining the versatility of data-independent acquisition (DIA) mass spectrometry (*11*, *12*) and the accuracy of ribosome profiling (*13*). We determined the protein number fractions $\psi_{p,i} \equiv [P_i]/[P]$ for >1900 proteins (labeled by i), with $[P] = \sum_i [P_i]$ being the total protein concentration, defined here as the number of proteins per cell volume. RNA sequencing (RNA-seq) was used to determine the mRNA number fractions $\psi_{m,i} \equiv [mR_i]/[mR]$ for the corresponding mRNAs, with $[mR] \equiv \sum_i [mR_i]$ being the total mRNA concentration (see materials and methods). In both cases, our data showed high reproducibility (fig. S3, A to C). The result for *E. coli* K-12 cells growing exponentially in glucose minimal medium is shown as a scatter plot of $\psi_{p,i}$ versus $\psi_{m,i}$ in Fig. 1B. We observed a strong correlation (r = 0.80) of mRNA and protein abundances along the diagonal (red line) across a vast range of abundances (10^{-2} to 10^{-6}). The histogram of $\psi_{p,i}/\psi_{m,i}$ is peaked around 1, with 50% of the genes within 1.7-fold (fig. S3E). We repeated the measurements for cells growing exponentially in three types of growth-limiting conditions in minimal medium (carbon limitation, anabolic limitation, and translational inhibition) (*14*, *15*), with growth rates ranging between 0.3 hour^{-1} and 0.9 hour^{-1}. A similar number of gene products were detected in these conditions, and the resulting scatter plots and histograms (fig. S3, F to S) looked very similar to those from bacteria grown in glucose minimal medium (fig. S3, D and E). These results, summarized in Fig. 1, C and D, indicate that the fractional abundances of mRNA and proteins are approximately the same. That is

$$\psi_{m,i} \approx \psi_{p,i} \qquad (1)$$

for most expressed genes in all growth conditions tested. The strong correlation between mRNA and protein fractions is also supported but less emphasized in several recent quantitative studies of *E. coli* protein expression (*13*, *16*–*18*) (fig. S3, T to W).

To probe how changes in protein and mRNA fractions relate to each other across growth conditions, we generated additional proteomics and transcriptomics datasets for more conditions under each type of growth limitation (fig. S4A) so that a smooth growth rate dependence could be obtained individually for the mRNA and protein fractions. We extrapolated these data to compute the fold change (FC) in the protein and mRNA fractions—$\mathrm{FC}(\psi_{p,i})$ and $\mathrm{FC}(\psi_{m,i})$, respectively—for each gene i (fig. S4B). The FC was calculated between the reference condition (wild-type cells grown in glucose minimal medium) and one with ~3× slower growth for each of the three types of growth limitation imposed. Their ratio, $\mathrm{FC}(\psi_{p,i})/\mathrm{FC}(\psi_{m,i})$, was even more tightly distributed than $\psi_{p,i}/\psi_{m,i}$ for each type of growth limitation (compare Fig. 1E with Fig. 1D), which indicated that the mRNA and protein fractions tightly covaried for most genes. The few exceptions that did not covary usually occurred in only one of the growth limitations and mostly corresponded to known targets of

[1]Department of Physics, University of California at San Diego, La Jolla, CA 92093, USA. [2]Departments of Medicine and Pharmacology, University of California at San Diego, La Jolla, CA 92093, USA. [3]Section of Molecular Biology, Division of Biological Sciences, University of California at San Diego, La Jolla, CA 92093, USA. [4]Faculty of Science, University of Zurich, Zürich, Switzerland. [5]Department of Biology, Institute of Molecular Systems Biology, ETH Zurich, Zürich, Switzerland. [6]Bavarian Center for Biomolecular Mass Spectrometry (BayBioMS), Technical University of Munich (TUM), Freising, Germany.
***Corresponding author. Email: hwa@ucsd.edu**
†These authors contributed equally to this work.

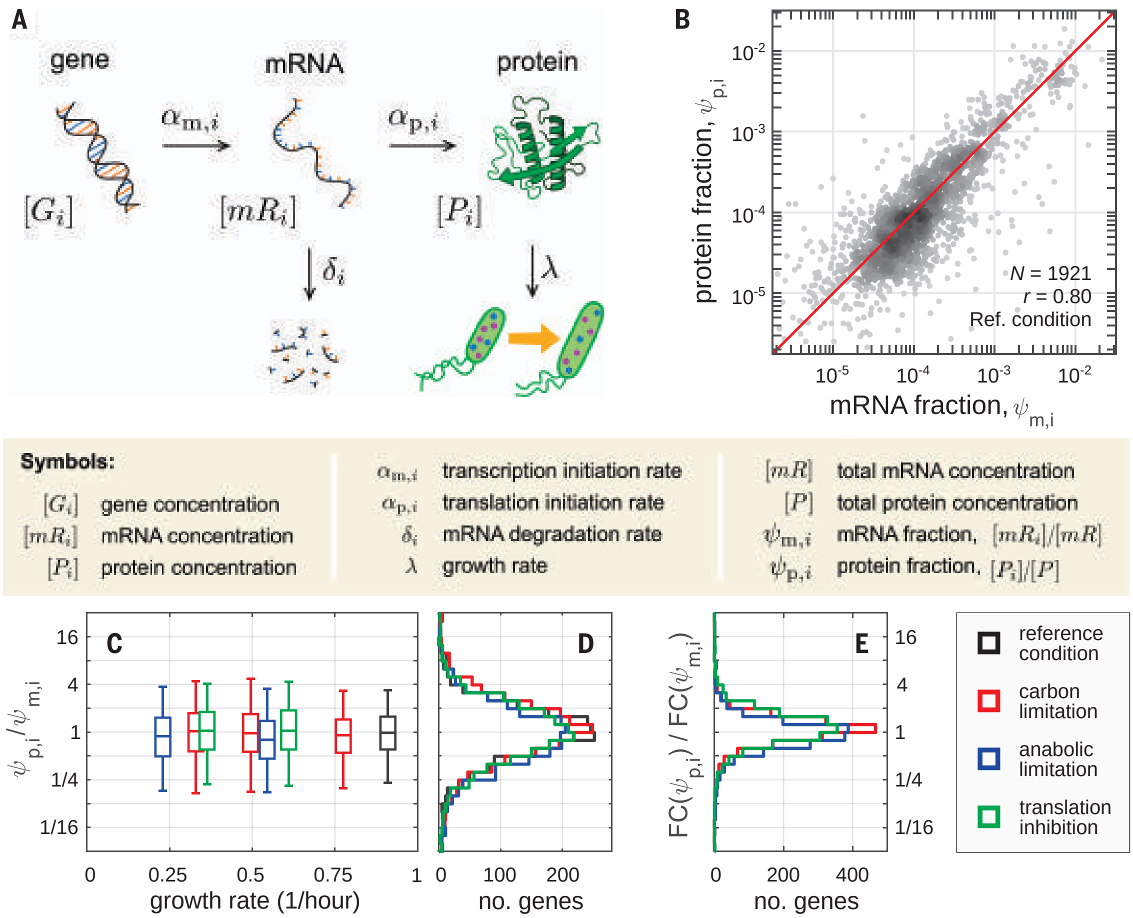


Fig. 1. Genome-wide mRNA and protein comparison. (**A**) Schematic illustration of the basic processes determining mRNA and protein concentrations in exponentially growing bacteria. The rate of each process can potentially vary across both genes and conditions; the symbols used throughout the study are described alongside the respective cellular processes (fig. S1). (**B**) For *E. coli* K-12 strain NCM3722 growing exponentially in glucose minimal medium (reference condition; growth rate, 0.91 hour^{-1}), the fractional number abundances of proteins ($\psi_{p,i}$, obtained from DIA/SWATH mass spectrometry) (*10*) and of mRNAs ($\psi_{m,i}$, obtained from RNA-seq; see materials and methods) for each gene *i* are shown as a scatter plot (number of genes and Pearson correlation coefficient *r* in the figure). The red line represents the diagonal, $\psi_{p,i} = \psi_{m,i}$. (**C**) The ratios of protein and mRNA fractions, $\psi_{p,i}/\psi_{m,i}$, are distributed around 1 for exponentially growing cultures under all growth conditions studied (fig. S3, E to S). These include the reference condition (black) as well as conditions of reduced growth achieved by limiting carbon catabolism (red), limiting anabolism (blue), or inhibiting translation (green) (materials and methods). The boxes and whiskers represent 50 and 90% of the genes, respectively; *x*-axis values give the corresponding growth rates. See tables S1 and S2 for lists of strains and conditions in this study and tables S3 and S4 for transcriptomics and proteomics data. (**D**) Distributions of the ratios $\psi_{p,i}/\psi_{m,i}$ obtained in reference condition and the slowest growing of each of the three types of limitations; same color code as in (C). The same plots also give the distributions of the relative translational initiation rate, $\alpha_{p,i}/\bar{\alpha}_p$ (see the text). (**E**) The FCs in protein and mRNA fractions for each gene *i* between the reference condition and the slowest growth condition, FC($\psi_{p,i}$) and FC($\psi_{m,i}$), were computed as described in fig. S4 for each one of the three growth limitations; the distribution of their ratio FC($\psi_{p,i}$)/FC($\psi_{m,i}$) is shown using the same color code as in (C). The histograms are narrowly distributed around 1, with more than half of the genes within 35% from the median. See table S5 for the FCs in translation efficiency for each gene.

posttranscriptional regulation (fig. S4, C to E, and table S5).

Total mRNA abundance matches translational capacity

From the steady-state relation between concentrations of individual mRNAs and proteins (fig. S1)

$$\alpha_{p,i}[mR_i] = \lambda[P_i] \quad (2)$$

where $\alpha_{p,i}$ is the translation initiation rate of each mRNA mR_i and λ denotes the growth rate, we can sum over contributions from all genes to obtain a relation between the flux of total protein synthesis and dilution

$$\bar{\alpha}_p[mR] = \lambda[P] \quad (3)$$

with $\bar{\alpha}_p \equiv \sum_i \alpha_{p,i}\psi_{m,i}$ being the average translational initiation rate (over all mRNAs). Because the ratio of Eq. 2 and Eq. 3 gives

$$\alpha_{p,i}/\bar{\alpha}_p = \psi_{p,i}/\psi_{m,i} \quad (4)$$

we see that Fig. 1D also provides the distribution of the relative translation initiation rates $(\alpha_{p,i}/\bar{\alpha}_p)$. The observed similarity between mRNA and protein fractions (Eq. 1) implies that the translation initiation rates $\alpha_{p,i}$ are similar for most mRNAs for each growth condition. Thus, the average translational initiation rate $\bar{\alpha}_p$ can be taken as representative most mRNAs.

Because the total protein concentration $[P]$ changes <10% across nutrient-limited conditions (fig. S2H), the total protein synthesis flux $\lambda[P]$ changes almost linearly with the growth rate λ. Consequently, Eq. 3 constrains the total mRNA concentration $[mR]$ and/or the average initiation rate $\bar{\alpha}_p$ to change with cellular growth. To understand how this constraint

is accommodated, we quantified the total mRNA amount by combining hybridization of ^{3}H-uracil–labeled RNA to genomic DNA with quantitative Northern blotting (fig. S5 and materials and methods). The result for carbon-limited growth (fig. S5H) was then converted to cellular concentration, i.e., $[mR]$ (supplementary text, note S1), and is shown as the red symbols in Fig. 2A (left vertical axis). These data allowed us to use Eq. 3 to obtain the average translation initiation rate $\bar{\alpha}_p$. The approximately linear growth rate dependence of $[mR]$ makes $\bar{\alpha}_p$ only weakly growth rate dependent (fig. S6A, left axis). The value of $\bar{\alpha}_p$ in turn enabled us to obtain the distribution of $\alpha_{p,i}$, the translational initiation rate of individual mRNAs, using Eq. 4 and the distributions of $\psi_{p,i}/\psi_{m,i}$ (Fig. 1D). The results for the reference and a slow, carbon-limiting growth condition exhibited weak dependence of $\alpha_{p,i}$ on both the mRNA species and growth condition (Fig. 2B and fig. S6B).

Constancy of ribosome spacing across mRNA and nutrient conditions

To understand how the relation between the total mRNA concentration and the total protein synthesis flux (Eq. 3) arises, we note that the total flux of peptide synthesis is given by $\lambda[P]\bar{\ell}_P$, where $\bar{\ell}_P$ is the average length of a protein—~250 amino acids (aa) across conditions (fig. S2G). Because genome-wide translation elongation rates are narrowly distributed (fig. S6C), this flux corresponds to the product of the concentration of actively translating ribosomes ($[Rb]_{act}$) and the speed of translational elongation (ε) as depicted in fig. S6D. That is

$$\lambda[P]\bar{\ell}_P = \varepsilon[Rb]_{act} \quad (5)$$

Both $[Rb]_{act}$ and ε have been characterized for a broad range of nutrient conditions (*7*). The active ribosome concentration is plotted in Fig. 2A (right vertical axis). The data exhibit a congruence with the total mRNA concentration, revealing a coordination of mRNA abundance and the cellular translational capacity, with an average spacing between translating ribosomes close to 200 nucleotides (nt) (Fig. 2C and fig. S6E). This translates to an average of number of ribosomes per mRNA $\bar{r} \approx 4$ for typical mRNAs 750 nt long. The observed proportionality between total mRNA and active ribosome concentrations implies that the average translational initiation rates $\bar{\alpha}_p$ and elongation rates ε are proportional (fig. S6A); this proportionality extends to individual mRNAs because of the similarity in protein and mRNA fractions observed above. By comparing the concentrations of protein and mRNAs (fig. S6, F and G), we confirmed that for most mRNA species, the ribosome spacing is clustered around 200 nt across conditions (Fig. 2, D and E, and fig. S6, H and I). The data are bounded by ~40 nt per ribosome, in accord with the physical packing limit (*19*).

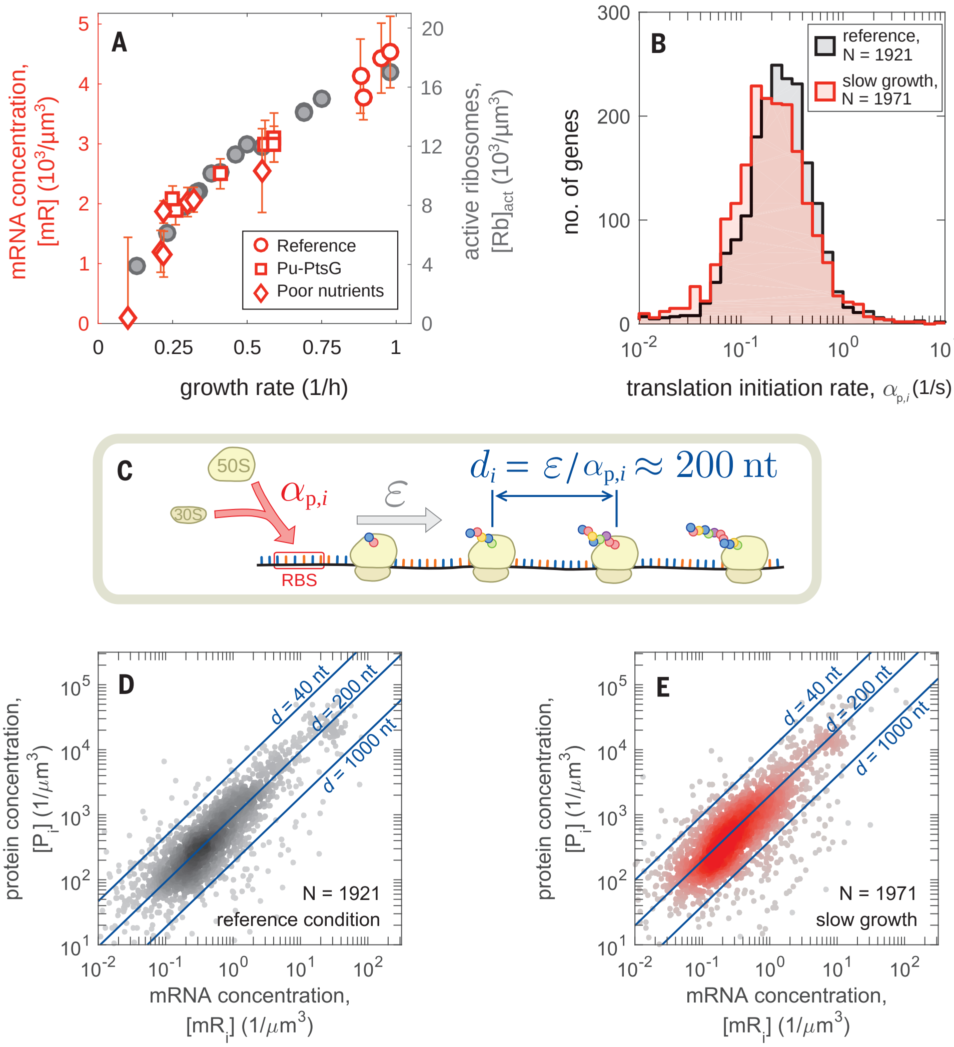


Fig. 2. Coordination of mRNA and ribosome abundances. (**A**) (Left axis; red symbols) Total concentration of mRNA is plotted against the growth rate. Total mRNA abundance and associated standard deviations are based on three measurements obtained as described in fig. S5 and the materials and methods. The measurements were performed for a range of growth conditions, including reference, glucose uptake titration (Pu-*ptsG*; table S1), and a host of poor-carbon sources. (Right axis; gray symbols) Concentration of active ribosomes in nutrient-limited conditions, converted from the data in (*7*) (reported per culture volume) using the total cellular volume shown in fig. S2, C to E. (**B**) Translation initiation rates, $\alpha_{p,i} = \bar{\alpha}_p\left(\psi_{p,i}/\psi_{m,i}\right)$, in reference (black) and carbon-limited (red) growth. (**C**) The spacing between consecutive translating ribosomes on an mRNA is given by the ratio between the ribosome elongation rate [similar across mRNAs; fig. S6 and (*7*)] and the translation initiation rate $\alpha_{p,i}$, which is also narrowly distributed [see (B)]. Our data give an average ribosome spacing of $d \approx 200$ nt (fig. S6D). RBS, ribosomal binding site. (**D**) Absolute mRNA and protein concentration for each gene in the reference condition, computed by combining the fractional abundances $\psi_{m,i}$ and $\psi_{p,i}$ with total mRNA abundances (A), total protein abundances, and cell volume (fig. S2 and supplementary text, note S1). Blue lines indicate the corresponding values of interribosome spacing d, calculated from the known elongation rates (~15.3 aa s^{-1}). (**E**) Same as (D), but for slow growth in the most carbon-limiting condition [growth rate, ~0.35 $hour^{-1}$; elongation rate, ~12.4 aa s^{-1} (*7*)].

mRNA degradation is largely condition independent

We investigated the mechanism behind the observed proportionality between the concentrations of total mRNA and active ribosomes (Fig. 2A). mRNA concentration is set by the balance between its synthesis and degradation (fig. S1). We performed kinetic experiments to determine mRNA degradation rates δ_i genome-wide in the reference and the slowest carbon-limiting condition by inhibiting transcription initiation with rifampicin and quantifying the relative mRNA concentrations at short time intervals by RNA-seq (fig. S7, A to C; materials and methods; and table S3). As an example, we measured time courses of changes in the relative mRNA concentrations of genes of the *nuo* operon in the two growth conditions (Fig. 3, A to C). The time course can be described as a delayed exponential decay, with the lag time reflecting the time needed for the RNA polymerase (RNAP) to reach the gene from the transcription start (fig. S7D) and the decay rate attributed to the turnover of that mRNA. This analysis yielded degradation rates for ~2700 mRNAs (materials and methods and table S6). Genome-wide, mRNA degradation rates were strongly correlated in the two growth conditions (fig. S7E). The average degradation rate was very similar (fig. S7F, vertical dashed lines), even after weighting by mRNA abundances (fig. S7G). In particular, the FC in δ_i is sharply peaked, with 90% of genes in the range of 0.50 to 1.57 (Fig. 3D). Overall, the data indicated a lack of dependence of degradation rates on either the mRNA species or the growth condition for most mRNAs, as observed in other studies (*20*, *21*). The cases where rates changed significantly could be associated to known posttranscriptional regulation (fig. S7, H and I).

Total mRNA synthesis flux is adjusted to match translational capacity

From the concentration and degradation rates of each mRNA species—$[mR_i]$ and δ_i, respectively—we can obtain the mRNA degradation flux, $\delta_i[mR_i]$, whose distributions are shown in Fig. 3E for the reference and slow growth conditions. A shift in the median of the two distributions is seen (vertical dashed lines), reflecting growth dependence of the total degradation flux, $\sum_i \delta_i[mR_i] = \bar{\delta}[mR]$, where $\bar{\delta} \equiv \sum_i \delta_i \psi_{\mathrm{m},i}$ is the average degradation rate across mRNAs. By the balance of mRNA synthesis and degradation in steady-state growth (fig. S1 and supplementary text, note S3), the total mRNA synthesis flux J_{mR} can be expressed as

$$J_{\mathrm{mR}} \equiv \sum_i \alpha_{\mathrm{m},i}[G_i] = \bar{\delta} \times [mR] \qquad (6)$$

Because the average degradation rate ($\bar{\delta}$) is affected little by growth conditions (fig. S7G), Eq. 6 predicts that the observed growth dependence of the total mRNA concentration $[mR]$ (Fig. 2A) is caused primarily by changes in mRNA synthesis flux J_{mR}.

We tested this prediction by directly measuring the total mRNA synthesis flux J_{mR} across the range of carbon-limited growth conditions, by pulse-labeling cultures with ^{3}H-uracil and hybridizing the labeled RNA to genomic DNA over short time intervals (fig. S8). These data showed a strong growth rate dependence (Fig. 3F, red symbols, left vertical axis), closely matching the observed growth dependence of the total mRNA concentration (reproduced as black symbols in Fig. 3F, right vertical axis). The total mRNA fluxes inferred from the degradation rates, $\bar{\delta}[mR]$ (orange crosses), were within 20% of the directly measured synthesis fluxes, which showcases the consistency of these two very different measurement approaches. Putting together the results in Figs. 2 and 3 shows that the global constraint (Eq. 3) is enforced primarily by matching total mRNA synthesis flux J_{mR} with translational capacity (Fig. 3G).

mRNA synthesis flux and transcriptional regulation

The synthesis flux of each mRNA species is given molecularly by the product of the transcription initiation rate per gene, $\alpha_{\mathrm{m},i}$, and the gene concentration, $[G_i]$ (Fig. 1A and fig. S1). The growth rate dependence of gene concentration is in turn given by the product of the number of chromosome replication origins (Ori) per cell volume, [Ori], and the gene dose relative to the Ori, $g_i \equiv [G_i]/[\mathrm{Ori}]$. Thus, the total mRNA synthesis flux can be expressed as

$$J_{\mathrm{mR}} = [\mathrm{Ori}] \times \sum_i \alpha_{\mathrm{m},i} g_i \qquad (7)$$

The relative gene dose g_i across growth rates can be obtained from the chromosomal position of the gene (*22*, *23*) and the chromosome replication time (fig. S9, A and B). Further including a weak growth rate dependence of the Ori concentration (*24*) (fig. S9, C to E), we obtain negative growth rate dependences for the concentration of genes $[G_i] = [\mathrm{Ori}] \times g_i$ at all chromosomal positions (Fig. 4A). It is then clear from Eq. 7 that the strong positive growth rate dependence seen for the total mRNA synthesis flux J_{mR} (Fig. 3F) cannot be accounted for by opposite growth rate dependences of gene concentrations and must involve systematic changes in the promoter activities $\alpha_{\mathrm{m},i}$. This was seen more explicitly by computing the distributions of the promoter activity $\alpha_{\mathrm{m},i}$, obtained for each gene using the known degradation fluxes $\delta_i[mR_i]$ and gene concentrations $[G_i]$ at steady state (eq. S3 in fig. S1; see also supplementary text, note S3, and data in table S6). The results (Fig. 4B) showed a broad range of promoter activity, spanning four orders of magnitude, with the high end (~0.3 s^{-1} in reference condition) approaching the maximum of ~1 s^{-1} given the transcriptional elongation speed of ~50 nt s^{-1} and a transcription elongation complex footprint of ~40 nt (*25*–*27*). A shift is seen between the distributions for reference (gray) and carbon-limited (red) growth conditions. In fact, for most genes, $\alpha_{\mathrm{m},i}$ in carbon-limited conditions decreases to about one-third of its value in the reference condition (Fig. 4C), a change that is comparable to the reduction in growth rate. Thus, the coordination of mRNA synthesis flux with the growth rate (Fig. 3F) is likely a result of genome-wide changes in transcription initiation between these conditions.

To look further into the determinants of transcription initiation, we used a canonical model of transcription regulation (Fig. 4D) (*28*, *29*), where the transcription initiation rate $\alpha_{\mathrm{m},i}$ for gene i is given by the product of the concentration of available RNA polymerases ($[\mathrm{RNAP}]_{\mathrm{av}}$) and the promoter on rate k_i. That is

$$\alpha_{\mathrm{m},i} = k_i \times [\mathrm{RNAP}]_{\mathrm{av}} \qquad (8)$$

where k_i captures the regulatory activities of all transcription factors acting on the promoter driving gene i (*28*, *29*). Using this expression for $\alpha_{\mathrm{m},i}$, the balance of mRNA synthesis and degradation (eq. S3 in fig. S1) can be written as

$$[\mathrm{RNAP}]_{\mathrm{av}}[\mathrm{Ori}] \times k_i g_i = \delta_i[mR_i] \qquad (9)$$

Quantitative relations connect transcriptional regulation to gene expression

From Eq. 9, we can derive two fundamental relations connecting transcription regulation to gene expression (supplementary text, note S4). Summing Eq. 9 over all genes, the balance of the total transcription flux becomes

$$[\mathrm{RNAP}]_{\mathrm{av}}[\mathrm{Ori}] \times \mathcal{K} = \bar{\delta} \times [mR] \qquad (10)$$

where $\mathcal{K} \equiv \sum_i k_i g_i$ describes the total on rate for promoters across the genome and is a measure of the total regulatory activity on transcription (weighted by gene dose). Given the proportionality between active ribosome and total mRNA concentrations, $\bar{r} = [Rb]_{\mathrm{act}}/[mR] \approx 4$ (Fig. 2A), Eq. 10 can be written as

$$[\mathrm{RNAP}]_{\mathrm{av}}[\mathrm{Ori}] \times \mathcal{K} = (\bar{\delta}/\bar{r}) \times [Rb]_{\mathrm{act}} \qquad (11)$$

This relation represents a fundamental constraint between the overall transcription activity $\left([\mathrm{RNAP}]_{\mathrm{av}}\mathcal{K}\right)$, the DNA content (via [Ori]), and the translational activity of the cell.

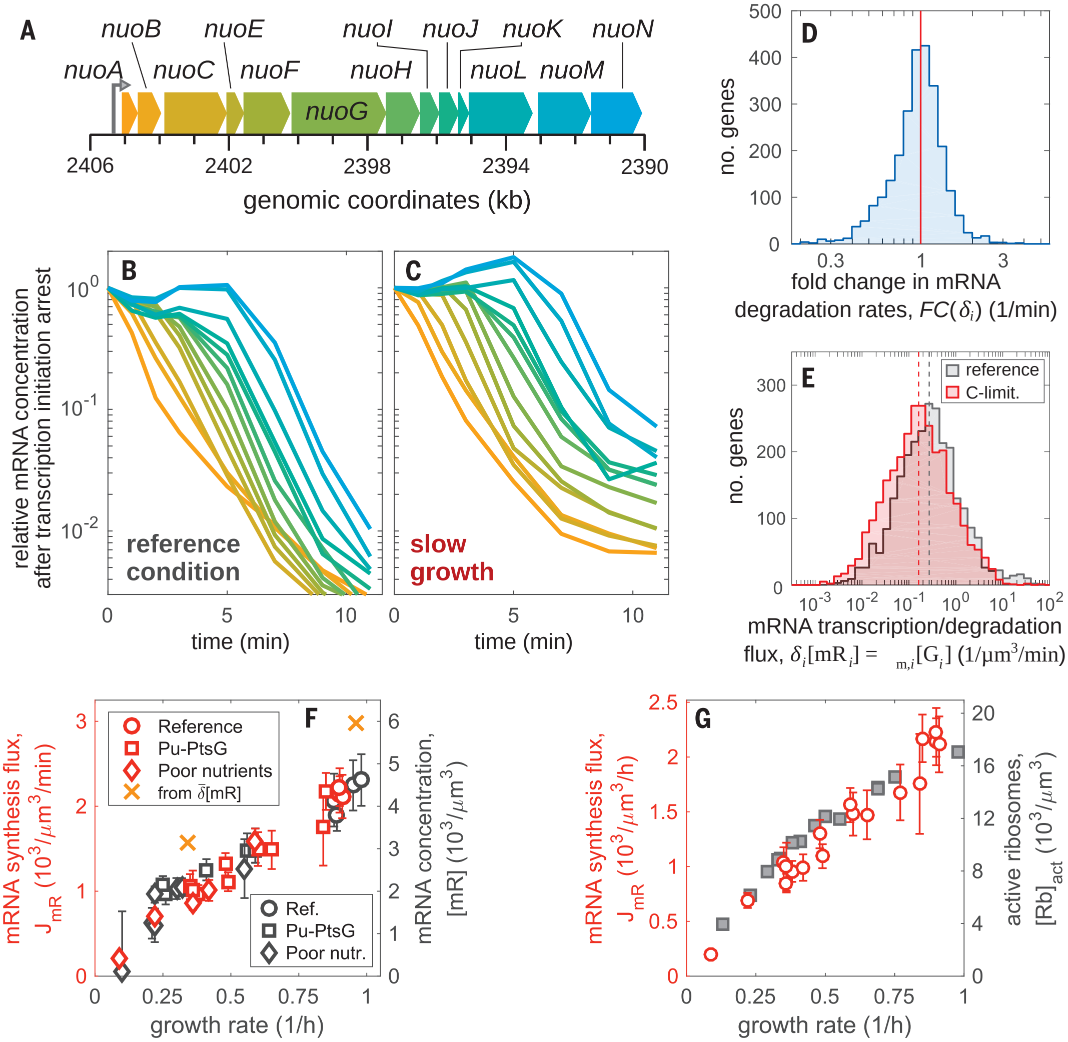


Fig. 3. mRNA degradation and synthesis. (**A** to **C**) Degradation of mRNA transcribed from the long *nuo* operon (A) in the reference condition (B) and carbon-limited condition (C). The abundance of mRNA was measured by RNA-seq over the course of 11 min after the blockage of transcription initiation by rifampicin (materials and methods and fig. S7). Although the abundance of the mRNA of genes proximal to the promoter (*nuoA*, orange) drops immediately after rifampicin treatment (at time t = 0), a lag is observed for genes progressively more distant from the promoter (from orange to blue). The lag time corresponds to the time elapsed between the transcription of the proximal and distant genes by RNAPs, which initiated transcription before the application of rifampicin (fig. S7D). (**D**) Histogram of FC of the mRNA degradation rates, FC(δ_i), between carbon-limited medium and reference condition for N = 2550 genes. Half of the FCs are within 25% from unity, and 90% of the FCs are in the range 0.50 to 1.57, implying that the degradation rates for most mRNAs do not change substantially between the reference and carbon-limited growth conditions. (**E**) Distribution of the mRNA degradation fluxes, $\delta_i[mR_i]$, computed from the mRNA concentration and degradation rates. These quantities should equate the mRNA synthesis fluxes, $\alpha_{m,i}[G_i]$, in steady-state conditions. Dashed lines indicate the median fluxes, 0.194 μm^{-3} min^{-1} in the reference condition and 0.108 μm^{-3} min^{-1} at slow growth. (**F**) (Left axis; red symbols) Total mRNA synthesis flux $J_{mR} = \sum_i \alpha_{m,i}[G_i]$ (transcripts synthesized per cell volume per unit time) for a variety of growth conditions, as indicated (see table S2 for growth conditions). The slope of radiolabel incorporated into mRNA over time was used to obtain the mRNA synthesis flux, whereas the error bars represent the standard deviation from six measurements at different time points after the label addition (fig. S10). The orange crosses indicate the total mRNA synthesis flux obtained from summing $\delta_i[mR_i]$ using the data in (E). (Right axis; black symbols) Absolute mRNA abundances (same data as Fig. 2A). (**G**) (Left axis; red symbols) Total RNA synthesis flux versus growth rate [same data as in (F)]. (Right axis; gray symbols) Concentration of active ribosomes (same data as Fig. 2A).

Another important relation can be obtained by taking the ratio of Eqs. 9 and 10. Noting that the mRNA degradation rates are closely distributed around the average $\bar{\delta}$ and are independent of growth conditions—i.e., $\delta_i \approx \bar{\delta}$ (fig. S7, F and G)—we obtain $k_i g_i/\mathcal{K} \approx [mR_i]/[mR] = \psi_{m,i}$. This relation extends further to the fractional protein abundances $\psi_{p,i} = [P_i]/[P]$ because of the established relation between protein and mRNA fractions (Eq. 1 and Fig. 1), which leads to

$$\frac{k_i g_i}{\mathcal{K}} \approx \psi_{m,i} \approx \psi_{p,i} \qquad (12)$$

This expression relates the (gene dose–weighted) regulatory activity on specific promoters ($k_i g_i$) to the mRNA and protein levels as determined by transcriptomics ($\psi_{m,i}$) and proteomics ($\psi_{p,i}$). Notably, $\psi_{p,i} = [P_i]/[P]$ gives approximately the cellular protein concentration $[P_i]$ because the total protein concentration varies only mildly with the growth rate, on average $[P] \approx 4.5 \times 10^6$ μm^{-3} (fig. S2H). Thus, Eq. 12 quantitatively connects regulatory activities at the promoter level (k_i) to cellular protein concentrations $[P_i]$ without explicit reference to the macroscopic machineries of gene expression. Equations 11 and 12 are the central quantitative results of this study. We suggest that Eq. 12 be viewed as a quantitative statement of the central dogma of bacterial gene expression, with Eq. 11 describing a system-level constraint on transcription and translation. In the following sections, we separately explore some consequences of these two central relations.

Global coupling in gene expression

According to Eq. 12, the mRNA and protein fractions corresponding to a given gene i are dependent not only on the regulatory activity on that gene, $k_i g_i$, but also on the total

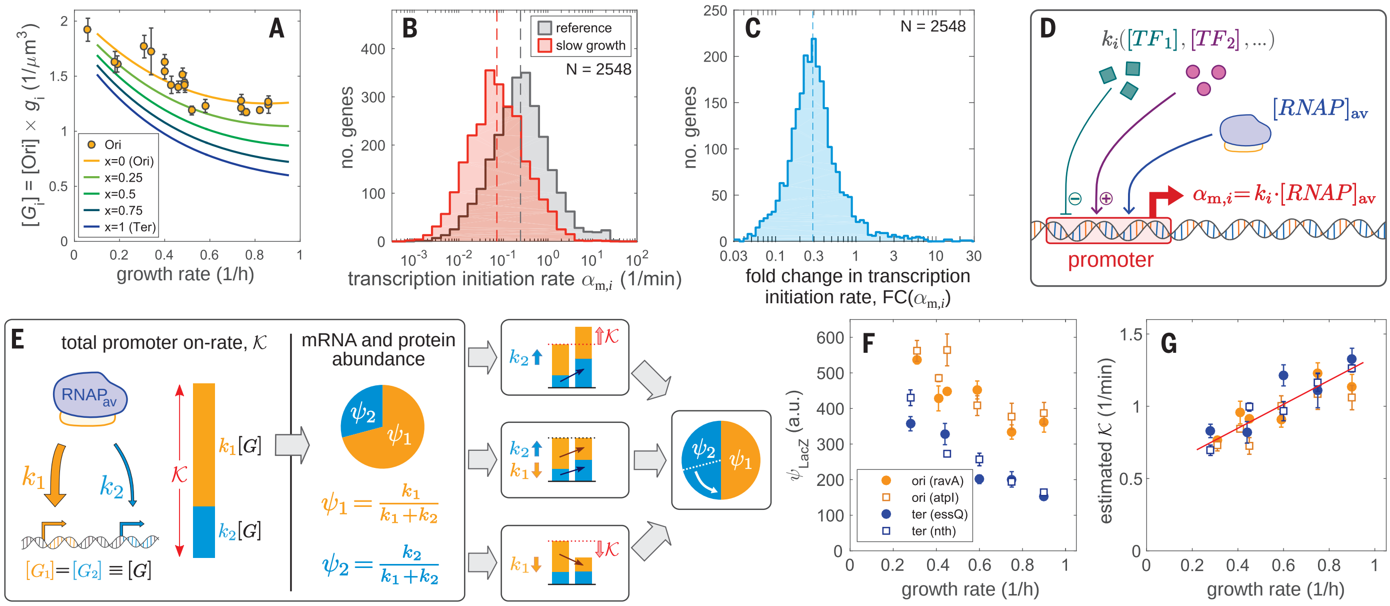


Fig. 4. Quantitative relations between promoter on rates, mRNA abundances, and protein abundances. (**A**) Growth rate dependence of gene concentration $[G_i]$ at various distances x from the origin of replication Ori (solid lines). These are computed as the product of the Ori concentration [Ori] [orange circles; shown in fig. S9C with raw data and standard errors from (*24*)] and the gene dose $g_i = [G_i]/[\mathrm{Ori}]$ (fig. S9B); see fig. S9 for details. (**B**) Distribution of transcription initiation rates $\alpha_{m,i}$ in reference condition (black) and slow growth (red), computed using the available mRNA abundances and degradation rates (see supplementary text, note S3, for details). Dashed lines indicate the median initiation rates in the two conditions (2.64 min^{-1} for reference condition, 0.87 min^{-1} for slow growth). (**C**) FC of the transcription initiation rates $\mathrm{FC}(\alpha_{m,i})$ between reference condition and slow growth. The data show a generalized decrease of initiation rates, with a median reduction of 0.29 (dashed line) at slow growth (λ = 0.3 hour^{-1}) compared with the reference condition (λ = 0.91 hour^{-1}). (**D**) Illustration of a canonical model of transcriptional regulation (*28*, *29*), with the transcription initiation rate for gene i, $\alpha_{m,i}$, depending on the promoter on rate k_i, which is modulated by transcription factors (TF_1, TF_2, ...), as well as on the cellular concentration of available RNAPs ($[\mathrm{RNAP}]_{av}$), as described by Eq. 8. (**E**) Cartoon illustrating the dependence of mRNA and protein abundances on the promoter on rates, as described by Eq. 12. Consider two genes with promoter on rates k_1 (orange) and k_2 (blue) and identical gene concentrations $[G_1] = [G_2 \equiv [G]$; the corresponding mRNA and protein fractions ($\psi_{m,1} = \psi_{p,1} \equiv \psi_1$ and $\psi_{m,2} = \psi_{p,2} \equiv \psi_2$, respectively) depend on both promoter on rates through the total regulatory activity $\mathcal{K} = (k_1 + k_2)[G]$ (in red). Three possible scenarios are illustrated. (Top) If k_2 increases while k_1 remains constant, then $\mathcal{K}$ increases, resulting in the reduction of protein and mRNA abundances for the orange gene despite it not being down-regulated at the transcriptional level. (Bottom) If only k_1 decreases while k_2 remains constant, then the proteins and mRNAs for the blue gene increase despite the lack of change at its promoter level. (Middle) If $\mathcal{K}$ is unchanged (because of compensating changes in k_1 and k_2 in this case), then the changes in protein and mRNA fractions would reflect changes at the regulatory level. (**F**) *E. coli* strains harboring constitutive expression of *lacZ* at various locations near *oriC* (orange) and near *terC* (blue; loci listed in the legend) were grown in carbon-limited conditions (see tables S1 and S2 for strains and conditions). LacZ protein abundance per culture volume (optical density × milliliter), obtained from the slopes of β-gal activity versus OD_{600} (optical density of a sample measured at a wavelength of 600 nm) (Miller units), is shown; error bars indicate standard errors from four measurements (materials and methods). a.u., arbitrary units. (**G**) The relative change in the total regulatory activity $\mathcal{K}$ across growth rates was estimated from the relative change in LacZ abundance using the data in (F) and Eq. 14. To do so, the LacZ abundance per culture volume was converted to protein fraction by dividing by total protein mass per culture volume (fig. S2F). The result shows a linear dependence of the total regulatory activity on the growth rate (red line). The absolute scale $\mathcal{K}$ was set for the reference condition using Eq. 10 with the values for the total mRNA synthesis flux J_{mR} obtained from Fig. 3F, the *oriC* concentration from Fig. 4A, and the available RNAP concentration estimated as described in the supplementary text, note S5.

regulatory activity, $\mathcal{K} \equiv \sum_j k_j g_j$. The latter dependence couples the expression of all genes in the cell, as illustrated in Fig. 4E. This dependence is explicitly seen when comparing FCs in gene expression across two different conditions

$$\mathrm{FC}([P_i]) = \mathrm{FC}(\psi_{p,i}) = \mathrm{FC}(k_i g_i)/\mathrm{FC}(\mathcal{K}) \quad (13)$$

In different growth conditions, where the promoter on rates k_i of many genes are affected, we generally expect the total rate $\mathcal{K}$ to vary—i.e., $\mathrm{FC}(\mathcal{K}) \neq 1$. Consequently, changes in the regulatory activity of a gene are generally expected to be different from the changes in the fractional abundances of the corresponding mRNA and protein. In fact, the latter might change even if the corresponding regulatory activity $k_i g_i$ is unchanged because of the overall change in regulatory activity $\mathcal{K}$ (illustrated in Fig. 4E).

To determine how the total regulatory activity $\mathcal{K}$ may change across growth conditions, we returned to the spectrum of carbon-limited growth conditions. The growth rate dependence of $\mathcal{K}$ can be deduced by applying the relation (Eq. 13) to constitutively expressed (i.e., unregulated) genes, for which k_i is constant. For this purpose, we inserted constitutively expressed *lacZ* genes at various locations x on the chromosome, with known gene doses $g(x)$ (Fig. 4A). The total rate $\mathcal{K}$ can then be obtained by measuring the concentration of LacZ, $[\mathrm{LacZ}(x)]$, for different growth rates as

$$\mathcal{K} \propto \frac{g(x)}{[\mathrm{LacZ}(x)]} \quad (14)$$

The data in carbon-limited conditions (Fig. 4F) show that $[\mathrm{LacZ}(x)]$ generally decreased at faster growth, with a steeper trend when *lacZ* was inserted near *ter* (blue) compared with when it was inserted near *ori* (orange). Upon calculating $\mathcal{K}$ using Eq. 14, the data collapsed on the same positive growth rate dependence (Fig. 4G). Although this set of experiments established the relative changes in $\mathcal{K}$ across

conditions, the absolute scale of $\mathcal{K}$ can be determined from Eq. 10 using the measured [Ori] (Fig. 4A) and the measured J_{mR} (Fig. 3F) for $\bar{\delta}[mR]$ (Eq. 5). As discussed in the supplementary text, note S5, the abundance of available RNAP, $[\mathrm{RNAP}]_{\mathrm{av}}$, can be estimated in the reference condition to be ~1300 $\mu\mathrm{m}^{-3}$, leading to $\mathcal{K}$ ~ 1.27 $\mu\mathrm{m}^3\ \mathrm{min}^{-1}$ in the reference condition.

Identifying promoter on rates

Knowledge of the magnitude of $\mathcal{K}$, together with the mRNA abundances and degradation rates, allowed us to compute the promoter on rates k_i for each gene i across growth conditions (supplementary text, note S4). The results for ~2500 genes (Fig. 5A) displayed a broad distribution across more than three orders of magnitude. Because k_i values were computed by combining several different genome-wide measurements (supplementary text, note 4), we sought to validate them by examining the on rates of genes belonging to the same operon to see whether the variation in the k_i of these cotranscribed genes are in fact small, as intuitively expected. The coefficients of variation in k_i within operons were not only much smaller than the values obtained by randomly extracting k_i from the different operons but were also significantly smaller than the variation in the concentrations of proteins associated to the same regulatory units (Fig. 5B and fig. S10A).

The complete set of gene expression rates generated in this work, including the promoter on rates, mRNA degradation rates, and translation initiation rates (table S6), allowed us to investigate, at the genome scale, the dominance of different factors controlling protein concentrations in *E. coli*, as well as their changes across conditions. By plotting the promoter on rates k_i against the protein concentrations in the reference condition (Fig. 5C, top), we see that the large range of protein concentrations can be largely attributed to differences in the promoter on rates, as opposed to other factors such as mRNA degradation rate, translation initiation rate, or the gene dose (other panels in Fig. 5C), which is in agreement with the simple scenario expressed by Eq. 12.

Proteins present at low concentrations tend to have lower translation initiation rates $\alpha_{\mathrm{p},i}$ and larger mRNA degradation rates δ_i compared with those at high concentrations (middle panels in Fig. 5C and fig. S10B); both effects tend to reduce the average number of proteins produced during the mRNA's lifetime (*3*) (fig. S10, C to F). When focusing on cotranscribed genes, we were also able to identify posttranscriptional effects (fig. S11, A and B) and evidence of premature transcriptional termination (fig. S11C). Still, these effects are rare and of small magnitude compared with the vast range of promoter on rates.

Unraveling the innate and regulatory effects on gene expression

FCs in protein abundance across conditions showed an almost perfect correlation with promoter on rates (Fig. 5D, top), whereas the effect of posttranscriptional regulation or gene copy number was negligible (other panels in Fig. 5D and fig. S10G). Thus, protein concentrations appear to be adjusted across conditions by modulating the promoter on rate, as described by Eq. 13. Yet, the typical changes in promoter on rates were very small (Fig. 5D), even though the full range of k_i spans more than three orders of magnitude (Fig. 5C and fig. S10, H and I). Based on proteomics data collected for a wide range of growth conditions from Mori *et al.* (*10*) (fig. S12, A to E), we find that that protein abundances for two-thirds of the genes vary by less than fivefold (fig. S12F) even though the absolute abundances vary by more than three orders of magnitude. Particularly small changes are observed for proteins engaged in housekeeping activities (fig. S12, G to I, and fig. S11, D to F) or those encoded by essential genes (fig. S12, F and J). Taken together, we conclude that innate promoter sequences determine the typical concentrations of most proteins, with transcriptional regulation providing fine adjustments depending on conditions. Some exceptions involve proteins needed in specific growth conditions (e.g., TCA proteins in aerobic conditions), whose genes are under strict transcriptional control (fig. S12, K to O).

Overall, promoter on rates for genes involved in protein synthesis increased with growth rate and largely accounted for the increase in the total promoter on rate K (Fig. 5E). Changes in k_i values for ribosomal proteins and elongation factors were similar, as expected, because they are largely cotranscribed (Fig. 5F, triangles, and fig. S13, A to D). Notably, they also matched the changes for the ribosomal RNA (rRNA) promoters (Fig. 5F, crosses). Thus, despite the known posttranscriptional regulation acting on ribosomal proteins (*30*), the coordination between ribosomal proteins and rRNAs is largely implemented by matching their promoter on rates. By contrast, the k_i values of other translation-affiliated proteins (e.g., initiation factors and tRNA synthetases) in different regulatory units present a variety of dependencies on the growth rate (fig. S13, E to H). Analogously, we found a variety of behaviors for the k_i of genes known to be controlled by cyclic adenosine 3′,5′-monophosphate (cAMP)–cAMP receptor protein (CRP) and typically expressed in carbon-poor conditions (*15*) (fig. S14). The strong increases in the protein levels of many catabolic proteins in carbon limitation are the result of a combination of faster promoter on rates and reduced total regulatory activity $\mathcal{K}$ at slow growth, which highlights the passive effects described in Fig. 4E.

Control of global mRNA synthesis by the anti-sigma factor Rsd

The constraint between transcription and translation (Eq. 11) must be somehow satisfied despite the observed changes in active ribosomes (Fig. 2A), DNA concentration (Fig. 4A), and total regulatory activity (Fig. 4G). The combination $[\mathrm{Ori}] \times \mathcal{K}$ has only a moderate dependence on growth rate (Fig. 6A). Instead, the concentration of available RNAP, estimated as the ratio of the mRNA synthesis flux (J_{mR}) and $[\mathrm{Ori}] \times \mathcal{K}$ based on Eqs. 6 and 10, exhibited a stronger growth rate dependence (Fig. 6B, left axis), approximately matching that of the concentration of active ribosomes (Fig. 6B, right axis).

A simple mechanism to change the availability of RNAPs is to change the abundance of the transcription machinery itself. However, our quantitative proteomics data showed that the cellular concentrations of RNAP components, including the housekeeping factor σ^{70} (encoded by the gene *rpoD*), are all constant across the growth rate range studied (Fig. 6C, squares). We checked expression levels of the two known modulators of σ^{70} function, 6*S* RNA (*31*, *32*) and the anti-σ^{70} protein Rsd (*33*–*35*). Although the concentration of 6*S* RNA is ~1/100 of that of σ^{70} (Fig. 6C, gray triangles) and thus is unlikely to affect the global transcription flux in these conditions, the Rsd concentration rose to that of σ^{70} as growth rate is reduced (Fig. 6C, red triangles). Thus, Rsd could be a regulator of global transcription by sequestering σ^{70} during exponential growth (Fig. 6D), even though it is thought to have its primary role in the stationary phase (*33*, *36*). We tested this scenario by characterizing the total mRNA synthesis flux in a Δ*rsd* strain. mRNA synthesis flux became nearly independent of growth rate (Fig. 6E, filled red circles), exceeding that of the wild-type strain (open red circles), especially at slow growth when Rsd is highly expressed. Without *rsd*, the synthesis flux was no longer matched to the translational capacity (filled symbols) in contrast to the tight matching observed in the wild-type strain (open symbols). Concomitantly, the Δ*rsd* strain exhibited a growth defect that was proportional to the amount of Rsd expression in wild-type cells in slow growth conditions where Rsd is expressed (Fig. 6F). Given the approximate constancy of mRNA turnover across growth conditions for wild-type cells (fig. S7, E to G), we propose that Rsd may have a central role in controlling total mRNA concentrations (Fig. 2A and Fig. 3F).

Discussion

We used comprehensive transcriptomic and proteomic studies, complemented by quantitative

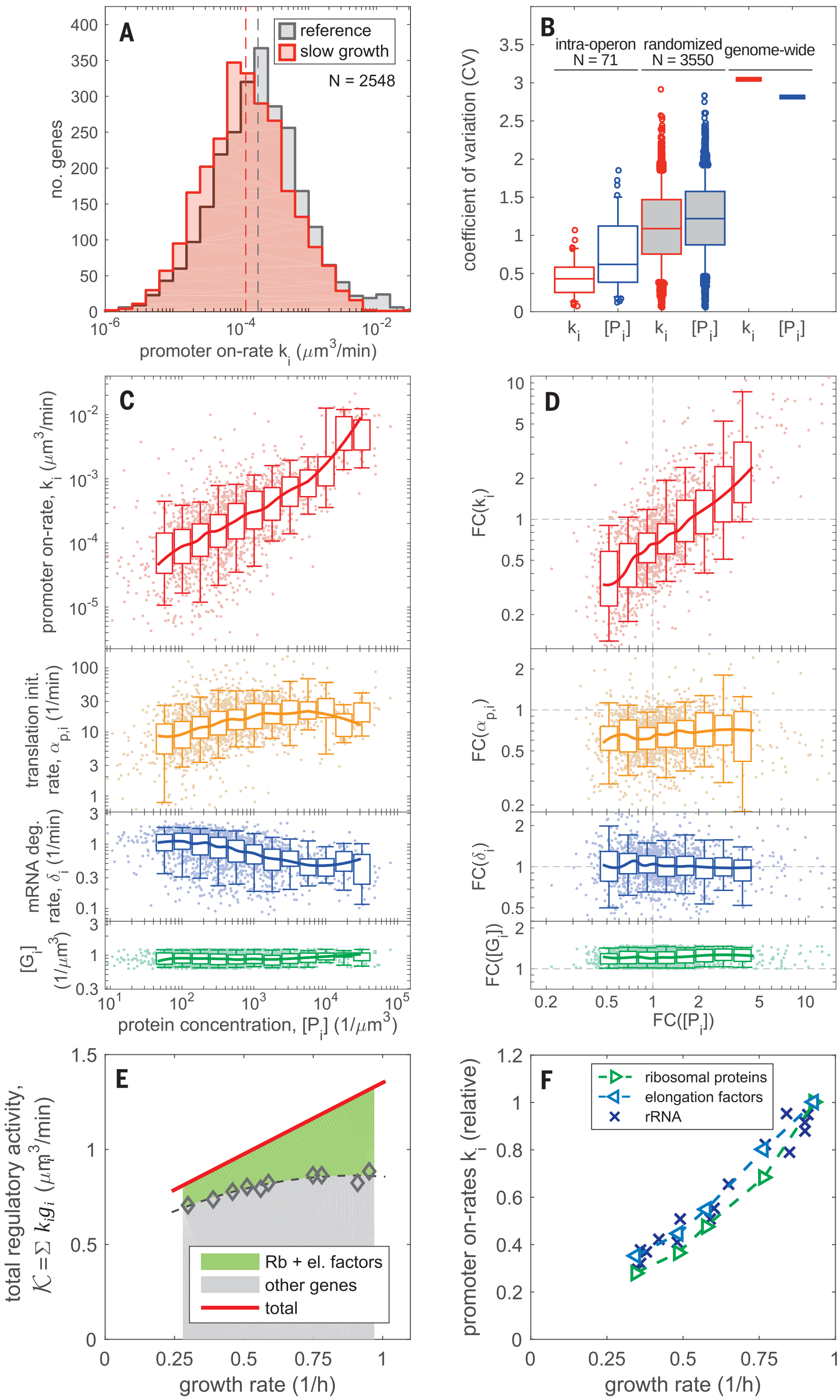


Fig. 5. Gene expression is primarily determined by the promoter on rates. (**A**) Distribution of promoter on rates k_i in the reference and slow growth conditions, obtained from the distribution of the translation initiation rate and the concentrations of available RNAP, $k_i = \alpha_{m,i}/[RNAP]_{av}$ (Eq. 8), as described in the supplementary text, note S4. The median promoter on rate (vertical dashed lines) shifts from 1.63×10^{-4} μm^3 min^{-1} in the reference condition ($\lambda \sim 0.9$ $hour^{-1}$) to 1.07×10^{-4} μm^3 min^{-1} in slow growth ($\lambda \sim 0.3$ $hour^{-1}$). This change is much less than the approximately threefold change in both the growth rate and the median transcription initiation rates (Fig. 4, B and C). (**B**) For 71 operons containing at least three genes as annotated in Ecocyc (*53*), we computed the coefficient of variation (CV) in the promoter on rates k_i or in the protein concentrations $[P_i]$ for genes within each operon in the reference condition. The average intraoperon CVs for the promoter on rates are significantly smaller than that computed for the protein concentrations $[P_i]$ ($P < 7 \times 10^{-7}$, unpaired *t* test) (fig. S10A). As a control, we randomly shuffled the genes across the operons 50 times, leading to sets of 3550 CVs (gray-filled boxes), and considered the CVs computed using all available genes (lines on the right). The CVs for the promoter on rates are also significantly smaller than those for all the other distributions ($P < 3 \times 10^{-35}$ when comparing with the randomized cases) and the genome-wide CVs. Boxes and whiskers indicate 50 and 90% intervals, respectively; median CVs are indicated by the central lines within the boxes. (**C**) Promoter on rates k_i, translation initiation rates $\alpha_{p,i}$, mRNA degradation rates δ_i, and gene concentrations $[G_i]$ are the four molecular parameters determining cellular concentration of a protein in a given growth condition (Fig. 1A; with the transcription initiation rate $\alpha_{m,i}$ given by k_i through Eq. 8). These four molecular parameters are plotted against the protein concentrations $[P_i]$ in the reference condition, binned according to the observed protein concentrations. Boxes and whiskers indicate 50 and 90% central intervals for the binned data; the solid lines represent moving averages. (**D**) Same as (C), but for the FCs of each quantity across growth conditions (slow growth compared with reference). All molecular parameters and concentrations shown in (A) to (D) are listed in table S6. (**E**) The sum of promoter on rates weighted by gene dose, $\mathcal{K} = \sum_i k_i g_i$ (red line; same as in Fig. 4G) is partitioned between the contribution from ribosomal proteins and translation elongation (Rb + el.) factors (green) and the rest of genes (gray area). Symbols indicate the partitioning obtained from the computed k_i across growth rates. The growth rate dependence of $\mathcal{K}$ largely stems from that of the promoter on rates of the translational genes. (**F**) Growth rate dependence of promoter on rates summed over different groups of genes: ribosomal proteins, elongation factors (encoded by *fusA*, *tufAB*, and *tsf*), and the rRNA operons. The activity of the rRNA operons was estimated from the synthesis flux of stable RNA (materials and methods and fig. S14).

measurement of total mRNA abundance and transcription flux, to determine the absolute mRNA and protein abundances, the mRNA degradation rates, and the promoter on rates for >1500 genes in *E. coli* for many growth conditions during steady-state growth (tables S3 to S5). The results revealed two simple rules on promoter and mRNA characteristics, which profoundly shape how *E. coli* responds to environmental changes while coping with global constraints: (i) Promoter on rates span more than three orders of magnitudes across genes but vary much less (at most approximately fivefold) across conditions for most genes. Thus, each gene is expressed within an innate abundance range across conditions—e.g., with ribosomal genes belonging to the most abundant and DNA replication proteins

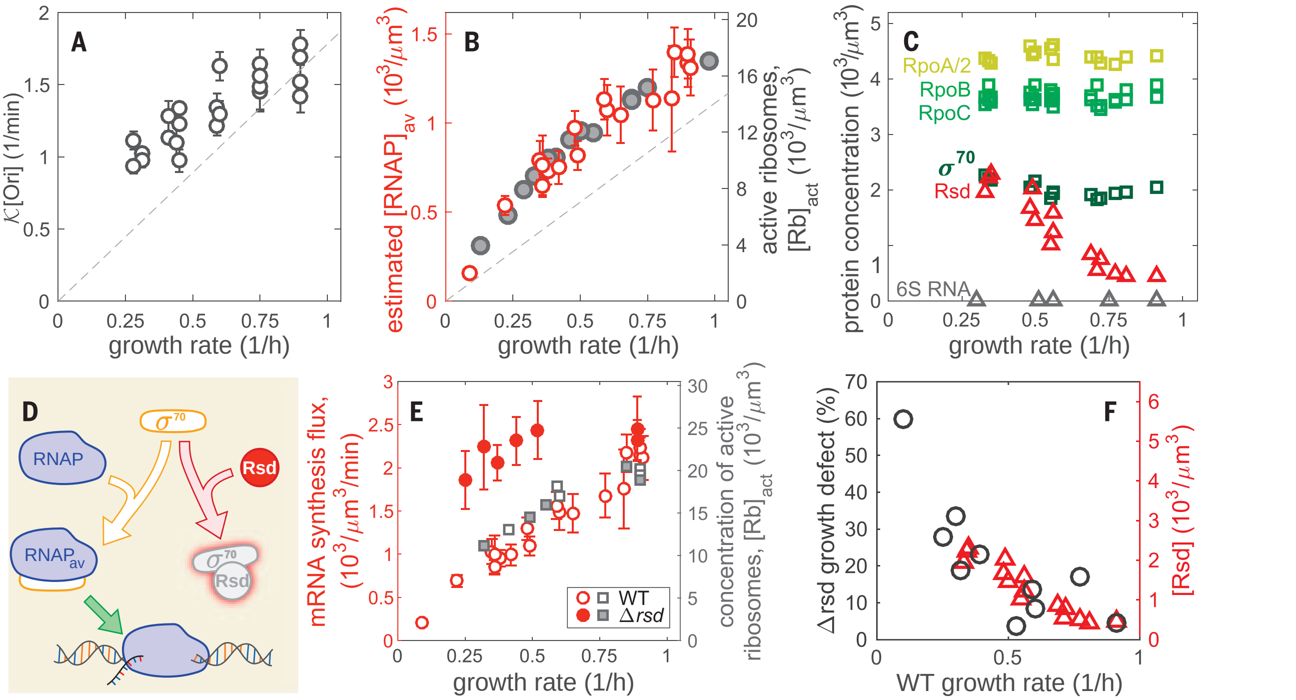


Fig. 6. The role of the anti-sigma factor Rsd in global regulation of mRNA synthesis. (**A**) Value of $\mathcal{K} \times$ [Ori] across growth rates, obtained from the values (data and standard errors) of the total regulatory activity $\mathcal{K}$ shown in Fig. 4G, multiplied by the interpolated values for [Ori] at the same growth rates (Fig. 4A). For comparison, the dashed line shows direct proportionality to the growth rate. (**B**) Concentration of available RNAPs (red symbols, left axis), estimated from the ratio between the measured mRNA synthesis flux (data and errors in Fig. 3G) and $\mathcal{K} \times$ [Ori] (using the interpolated curves in Fig. 4, A and G). This quantity shows a stronger dependence on the growth rate compared with $\mathcal{K} \times$ [Ori] in (A) and has the same growth rate dependence as the concentration of active ribosomes (gray symbols, right axis). (**C**) The concentrations of various components of the transcription machinery in carbon-limited conditions are plotted against the growth rate. Components of the core enzyme, RpoABC, and the major sigma factor σ^{70} are shown as squares. Known modulators of σ^{70}, Rsd and 6S RNA, are shown as triangles. The protein concentrations are determined from mass spectrometry (*10*), whereas the concentration of 6S RNA is determined from RNA-seq and the concentration of total mRNA concentration (fig. S4). (**D**) Cartoon illustrating the control of RNAP availability through the known σ^{70}-sequestration function of Rsd (*33*, *54*). (**E**) Comparison of mRNA synthesis fluxes between wild-type (WT) (open symbols) and Δ*rsd* strain (filled symbols). (Left axis) Total mRNA synthesis flux of Δ*rsd* strain (red filled circles) and wild-type (red open circles); standard errors are computed as in Fig. 3F. (Right axis) Concentration of active ribosomes computed from the measured total RNA for the two strains and the fraction of active ribosomes observed in carbon-limited growth (*7*). (**F**) The growth defect of the Δ*rsd* strain, defined as the percent reduction in growth rate compared with that of wild-type cells in the same growth condition (black circles, left axis), is plotted against the growth rate of wild-type cells for the range of carbon-limited growth conditions. The observed growth reduction matches Rsd expression of wild-type cells in the same conditions [red triangles, right axis; same data as in (C)].

belonging to one of the least abundant classes. (ii) mRNA characteristics, including translation initiation rate and mRNA degradation rate, vary little (less than twofold for half of the genes) across genes and conditions. The translation initiation rates are sufficiently rapid to maintain a high density of ribosomes on the mRNA (five ribosomes per kilobase; Fig. 2A and fig. S6E), resulting in high protein production despite short mRNA half-lives. The rules governing promoter and mRNA characteristics deduced here dictate, to a large extent, *E. coli*'s strategy to implement gene regulation while complying with the constraints on total protein concentration and a limited translation capacity. This can be cast into two design principles of gene expression, as summarized below.

Global coordination between transcription and translation

Concentrations of translating ribosomes are known to increase linearly with the growth rate of the bacteria (Fig. 2A). Because of the constant density of translating ribosomes on most mRNA, the total mRNA concentration must also scale similarly. As the total mRNA pool is specified by the total mRNA synthesis rate (given the constant mRNA turnover rate across conditions), total mRNA synthesis is balanced with translating ribosomes concentration—a crucial condition captured by Eq. 11. We refer to this balance as the principle of transcription-translation coordination. We showed that *E. coli* implements this coordination across nutrient conditions primarily by adjusting the available RNAP concentration through the anti-σ^{70} factor Rsd (Fig. 6, D to F).

If this coordination is broken, as in the case of the Δ*rsd* mutant (Fig. 6E), then rule 2 cannot hold as long as the constraints on translation capacity and protein density hold. An oversupply of mRNA with respect to ribosomes is expected to decrease the rate of translation initiation (because of competition for limited ribosomes) and increase the rate of mRNA degradation [because of reduced protection of mRNA by elongation ribosomes against ribonuclease (RNase) activity] (*37*, *38*). Aside from the futile cycle involving the synthesis and degradation of mRNAs and affecting growth (Fig. 6F), breaking rule 2 would complicate the otherwise simple relation between transcriptional regulation and protein concentrations of the wild-type system.

The predominant role of transcriptional control in setting protein concentrations

The similarity of mRNA characteristics (rule 2) together with the vast disparity of promoter characteristics (rule 1) across genes in a given condition implies that protein abundances are predominantly set by the promoter characteristics, specifically, the promoter on rates. Furthermore, because mRNA characteristics remain similar across different growth conditions (rule 2), changes in protein concentrations

across conditions must arise primarily from changes in the promoter on rates—i.e., through transcriptional regulation (Fig. 5D). We refer to this strong effect of transcription on gene expression as the principle of transcriptional predominance.

The strong mRNA-protein correlations are at odds with early studies on bacterial gene expression (*39*–*45*). These differences may originate in part from technological advances in transcriptomics and proteomics over the years (*10*, *45*, *46*). Further, given the very different turnover rates of proteins and mRNAs, accurate mRNA-protein comparison requires ensuring steady-state conditions, which was a crucial component of our experimental design (materials and methods). We note that strong mRNA-protein correlation, similar to what we report here, was also contained in a number of recent datasets (*13*, *16*–*18*) (fig. S3, T to W). However, such comparisons were not articulated as the main messages in these studies because their focus was on the variability in translational characteristics among mRNAs. Although we also see such differences (fig. S4, C to E), our data [as well as those from (*14*–*17*)] show that such variabilities do not represent the general behavior of most genes.

Although setting protein concentrations transcriptionally appears simple, a quantitative relation between promoter on rates and protein concentrations is complicated by the fact that the total protein concentration does not vary by much, despite strong changes in total transcriptional activity. Because protein concentrations do not depend on the absolute flux of the corresponding mRNAs but rather on the fraction of the total transcription flux, a hypothetical doubling of all promoter on rates, as we alluded to above, will have no effect on mRNA and protein concentrations. Furthermore, the relations between promoter on rates and protein concentrations are insulated from growth-dependent differences in shared machinery, such as RNAPs and ribosomes.

Global transcriptional coupling and its consequences

Because the protein output of a given promoter depends on the total regulatory activity $\mathcal{K}$ in Eq. 12, nonintuitive relations between promoters and protein concentration can arise when $\mathcal{K}$ changes across conditions. The latter is likely whenever there is a substantial change in growth conditions—e.g., because of changes in the on rates of ribosomal genes (fig. S12H and Fig. 5H). Hence, it is generally incorrect to infer regulatory activities directly from changes in mRNA and protein concentrations (Fig. 4E). This effect of global coupling that reflects the passive regulatory effect would hardly affect the result of most classical studies, which typically have involved large changes in the output of a few individual promoters. However, for most genes whose expression changed moderately across conditions (e.g., less than twofold for 60% of the genes under carbon-limited growth; see Fig. 5D), the effect of global coupling would be substantial.

Our work provides a quantitative framework to distinguish gene-specific regulatory effects from global interactions in gene expression studies. Knowing the promoter on rates k_i for individual genes facilitates a direct, promoter-centric view of regulation across conditions at the genome scale (Fig. 5). This improves upon previous estimates of promoter activity based on protein synthesis fluxes (*47*), which mix systemic effects, such as RNAP availability, with gene-specific regulatory effects (supplementary text, note 4). These results would thus be of use in deciphering the behaviors of endogenous genetic circuits and in guiding the design of synthetic circuits in variable growth conditions (*48*–*50*).

The results described here are specific to bacteria. Eukaryotic gene expression involves complex posttranscriptional regulation, including protein secretion (*51*) and degradation through ubiquination and autophagy (*52*). Global constraints are less understood, in particular the extent to which protein density may vary across conditions. Even quantifying the cell volume may be difficult because large portions within a cell may be occupied by subcellular compartments (e.g., vacuoles) that do not contribute to the cytosol. Nonetheless, our study provides a framework to quantitatively explore gene expression in such complex systems.

Materials and methods summary

Experimental methods for cellular growth; RNA-seq; quantification of mRNA abundance, synthesis fluxes, and degradation rates; and numerical and statistical methods are reported in the supplementary materials.

REFERENCES AND NOTES

1. A. L. Goldberg, A. C. St. John, Intracellular protein degradation in mammalian and bacterial cells: Part 2. *Annu. Rev. Biochem.* **45**, 747–803 (1976). doi: 10.1146/annurev.bi.45.070176.003531; pmid: 786161
2. K. Nath, A. L. Koch, Protein degradation in *Escherichia coli*: II. Strain differences in the degradation of protein and nucleic acid resulting from starvation. *J. Biol. Chem.* **246**, 6956–6967 (1971). doi: 10.1016/S0021-9258(19)45938-3; pmid: 4942328
3. J. Paulsson, Models of stochastic gene expression. *Phys. Life Rev.* **2**, 157–175 (2005). doi: 10.1016/j.plrev.2005.03.003
4. S. Klumpp, Z. Zhang, T. Hwa, Growth rate-dependent global effects on gene expression in bacteria. *Cell* **139**, 1366–1375 (2009). doi: 10.1016/j.cell.2009.12.001; pmid: 20064380
5. J. Lin, A. Amir, Homeostasis of protein and mRNA concentrations in growing cells. *Nat. Commun.* **9**, 4496 (2018). doi: 10.1038/s41467-018-06714-z; pmid: 30374016
6. M. Scott, C. W. Gunderson, E. M. Mateescu, Z. Zhang, T. Hwa, Interdependence of Cell Growth and Gene Expression: Origins and Consequences. *Science* **330**, 1099–1102 (2010). doi: 10.1126/science.1192588; pmid: 21097934
7. X. Dai *et al*., Reduction of translating ribosomes enables *Escherichia coli* to maintain elongation rates during slow growth. *Nat. Microbiol.* **2**, 16231 (2016). doi: 10.1038/nmicrobiol.2016.231; pmid: 27941827
8. R. Milo, What is the total number of protein molecules per cell volume? A call to rethink some published values. *BioEssays* **35**, 1050–1055 (2013). doi: 10.1002/bies.201300066; pmid: 24114984
9. E. R. Oldewurtel, Y. Kitahara, S. van Teeffelen, Robust surface-to-mass coupling and turgor-dependent cell width determine bacterial dry-mass density. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2021416118 (2021). doi: 10.1073/pnas.2021416118; pmid: 34341116
10. M. Mori *et al*., From coarse to fine: The absolute *Escherichia coli* proteome under diverse growth conditions. *Mol. Syst. Biol.* **17**, e9536 (2021). doi: 10.15252/msb.20209536; pmid: 34032011
11. L. C. Gillet *et al*., Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: A new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics* **11**, 016717 (2012). doi: 10.1074/mcp.O111.016717; pmid: 22261725
12. C. Ludwig *et al*., Data-independent acquisition-based SWATH-MS for quantitative proteomics: A tutorial. *Mol. Syst. Biol.* **14**, e8126 (2018). doi: 10.15252/msb.20178126; pmid: 30104418
13. G.-W. Li, D. Burkhardt, C. Gross, J. S. Weissman, Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell* **157**, 624–635 (2014). doi: 10.1016/j.cell.2014.02.033; pmid: 24766808
14. C. You *et al*., Coordination of bacterial proteome with metabolism by cyclic AMP signalling. *Nature* **500**, 301–306 (2013). doi: 10.1038/nature12446
15. S. Hui *et al*., Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria. *Mol. Syst. Biol.* **11**, 784 (2015). doi: 10.15252/msb.20145697; pmid: 25678603
16. Y. Taniguchi *et al*., Quantifying *E. coli* Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells. *Science* **329**, 533–538 (2010). doi: 10.1126/science.1188308; pmid: 20671182
17. C. Del Campo, A. Bartholomäus, I. Fedyunin, Z. Ignatova, Secondary Structure across the Bacterial Transcriptome Reveals Versatile Roles in mRNA Regulation and Function. *PLOS Genet.* **11**, e1005613 (2015). doi: 10.1371/journal.pgen.1005613; pmid: 26495981
18. D. Choe *et al*., Adaptive laboratory evolution of a genome-reduced *Escherichia coli*. *Nat. Commun.* **10**, 935 (2019). doi: 10.1038/s41467-019-08888-6; pmid: 30804335
19. F. Mohammad, R. Green, A. R. Buskirk, A systematically-revised ribosome profiling method for bacteria reveals pauses at single-codon resolution. *eLife* **8**, e42591 (2019). doi: 10.7554/eLife.42591.001; pmid: 30724162
20. J. A. Bernstein, A. B. Khodursky, P.-H. Lin, S. Lin-Chao, S. N. Cohen, Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. *PNAS* **99**, 9697–9702 (2002). doi: 10.1073/pnas.112318199
21. R. L. Coffman, T. E. Norris, A. L. Koch, Chain elongation rate of messenger and polypeptides in slowly growing *Escherichia coli*. *J. Mol. Biol.* **60**, 1–11 (1971). doi: 10.1016/0022-2836(71)90442-6; pmid: 4937191
22. H. Bremer, G. Churchward, An examination of the Cooper-Helmstetter theory of DNA replication in bacteria and its underlying assumptions. *J. Theor. Biol.* **69**, 645–654 (1977). doi: 10.1016/0022-5193(77)90373-3; pmid: 607026
23. S. Cooper, C. E. Helmstetter, Chromosome replication and the division cycle of *Escherichia coli* Br. *J. Mol. Biol.* **31**, 519–540 (1968). doi: 10.1016/0022-2836(68)90425-7; pmid: 4866337
24. H. Zheng *et al*., General quantitative relations linking cell growth and the cell cycle in *Escherichia coli*. *Nat. Microbiol.* **5**, 995–1001 (2020). doi: 10.1038/s41564-020-0717-x; pmid: 32424336
25. B. Krummel, M. J. Chamberlin, Structural analysis of ternary complexes of *Escherichia coli* RNA polymerase: Deoxyribonuclease I footprinting of defined complexes. *J. Mol. Biol.* **225**, 239–250 (1992). doi: 10.1016/0022-2836(92)90918-A; pmid: 1593619
26. B. Z. Ring, W. S. Yarnell, J. W. Roberts, Function of E. coli RNA polymerase σ factor-σ^{70} in promoter-proximal pausing. *Cell* **86**, 485–493 (1996). doi: 10.1016/S0092-8674(00)80121-X; pmid: 8756730
27. M. Zhu, M. Mori, T. Hwa, X. Dai, Disruption of transcription-translation coordination in *Escherichia coli* leads to

premature transcriptional termination. *Nat. Microbiol.* **4**, 2347–2356 (2019). doi: 10.1038/s41564-019-0543-1; pmid: 31451774
28. L. Bintu *et al*., Transcriptional regulation by the numbers: Models. *Curr. Opin. Genet. Dev.* **15**, 116–124 (2005). doi: 10.1016/j.gde.2005.02.007; pmid: 15797194
29. L. Bintu *et al*., Transcriptional regulation by the numbers: Applications. *Curr. Opin. Genet. Dev.* **15**, 125–135 (2005). doi: 10.1016/j.gde.2005.02.006; pmid: 15797195
30. J. L. Yates, A. E. Arfsten, M. Nomura, In vitro expression of Escherichia coli ribosomal protein genes: Autogenous inhibition of translation. *Proc. Natl. Acad. Sci. U.S.A.* **77**, 1837–1841 (1980). doi: 10.1073/pnas.77.4.1837; pmid: 6445562
31. K. M. Wassarman, G. Storz, 6S RNA regulates *E. coli* RNA polymerase activity. *Cell* **101**, 613–623 (2000). doi: 10.1016/S0092-8674(00)80873-9; pmid: 10892648
32. K. M. Wassarman, 6S RNA, a global regulator of transcription. *Microbiol. Spectr.* **6**, 6.3.06 (2018). doi: 10.1128/microbiolspec.RWR-0019-2018
33. M. Jishage, A. Ishihama, A stationary phase protein in *Escherichia coli* with binding activity to the major σ subunit of RNA polymerase. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 4953–4958 (1998). doi: 10.1073/pnas.95.9.4953
34. T. M. Gruber, C. A. Gross, Multiple sigma subunits and the partitioning of bacterial transcription space. *Annu. Rev. Microbiol.* **57**, 441–466 (2003). doi: 10.1146/annurev.micro.57.030502.090913; pmid: 14527287
35. A. Lal, S. Krishna, A. Sai, N. Seshasayee, Regulation of Global Transcription in *Escherichia coli* by Rsd and 6S RNA. *Genes Genom. Genet.* **8**, 2079–2089 (2018). doi: 10.1534/g3.118.200265; pmid: 29686109
36. S. E. Piper, J. E. Mitchell, D. J. Lee, S. J. W. Busby, A global view of *Escherichia coli* Rsd protein and its interactions. *Mol. Biosyst.* **5**, 1943–1947 (2009). doi: 10.1039/b904955j; pmid: 19763331
37. A. Deana, J. G. Belasco, Lost in translation: The influence of ribosomes on bacterial mRNA decay. *Genes Dev.* **19**, 2526–2533 (2005). doi: 10.1101/gad.1348805; pmid: 16264189
38. M. P. Hui, P. L. Foley, J. G. Belasco, Messenger RNA Degradation in Bacterial Cells. *Annu. Rev. Genet.* **48**, 537–559 (2014). doi: 10.1146/annurev-genet-120213-092340; pmid: 25292357
39. L. Nie, G. Wu, W. Zhang, Correlation of mRNA expression and protein abundance affected by multiple sequence features related to translational efficiency in *Desulfovibrio vulgaris*: A quantitative analysis. *Genetics* **174**, 2229–2243 (2006). doi: 10.1534/genetics.106.065862; pmid: 17028312
40. P. Lu, C. Vogel, R. Wang, X. Yao, E. M. Marcotte, Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat. Biotechnol.* **25**, 117–124 (2007). doi: 10.1038/nbt1270; pmid: 17187058
41. R. de Sousa Abreu, L. O. Penalva, E. M. Marcotte, C. Vogel, Global signatures of protein and mRNA expression levels. *Mol. Biosyst.* **5**, 1512–1526 (2009). doi: 10.1039/b908315d; pmid: 20023718
42. J. M. Laurent *et al*., Protein abundances are more conserved than mRNA abundances across diverse taxa. *Proteomics* **10**, 4209–4212 (2010). doi: 10.1002/pmic.201000327; pmid: 21089048
43. T. Maier *et al*., Quantification of mRNA and protein and integration with protein turnover in a bacterium. *Mol. Syst. Biol.* **7**, 511 (2011). doi: 10.1038/msb.2011.38; pmid: 21772259
44. U. Omasits *et al*., Directed shotgun proteomics guided by saturated RNA-seq identifies a complete expressed prokaryotic proteome. *Genome Res.* **23**, 1916–1927 (2013). doi: 10.1101/gr.151035.112; pmid: 23878158
45. Y. Liu, A. Beyer, R. Aebersold, On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell* **165**, 535–550 (2016). doi: 10.1016/j.cell.2016.03.014; pmid: 27104977
46. F. Ozsolak, P. M. Milos, RNA sequencing: Advances, challenges and opportunities. *Nat. Rev. Genet.* **12**, 87–98 (2011). doi: 10.1038/nrg2934; pmid: 21191423
47. L. Gerosa, K. Kochanowski, M. Heinemann, U. Sauer, Dissecting specific and global transcriptional regulation of bacterial gene expression. *Mol. Syst. Biol.* **9**, 658 (2013). doi: 10.1038/msb.2013.14; pmid: 23591774
48. J. A. N. Brophy, C. A. Voigt, Principles of genetic circuit design. *Nat. Methods* **11**, 508–520 (2014). doi: 10.1038/nmeth.2926; pmid: 24781324
49. K. Michalodimitrakis, M. Isalan, Engineering prokaryotic gene circuits. *FEMS Microbiol. Rev.* **33**, 27–37 (2009). doi: 10.1111/j.1574-6976.2008.00139.x; pmid: 19016883
50. N. Nandagopal, M. B. Elowitz, Synthetic biology: Integrated gene circuits. *Science* **333**, 1244–1248 (2011). doi: 10.1126/science.1207084; pmid: 21885772
51. Z. Sun, J. L. Brodsky, Protein quality control in the secretory pathway. *J. Cell Biol.* **218**, 3171–3187 (2019). doi: 10.1083/jcb.201906047; pmid: 31537714
52. I. Dikic, Proteasomal and autophagic degradation systems. *Annu. Rev. Biochem.* **86**, 193–224 (2017). doi: 10.1146/annurev-biochem-061516-044908; pmid: 28460188
53. P. D. Karp *et al*., The EcoCyc Database. *Ecosal Plus* **8**, 10.1128/ecosalplus.ESP-0006-2018 (2018). doi: 10.1128/ecosalplus.ESP-0006-2018; pmid: 30406744
54. L. F. Westblade *et al*., Studies of the *Escherichia coli* Rsd-σ^{70} complex. *J. Mol. Biol.* **335**, 685–692 (2004). doi: 10.1016/j.jmb.2003.11.004; pmid: 14687566

ACKNOWLEDGMENTS

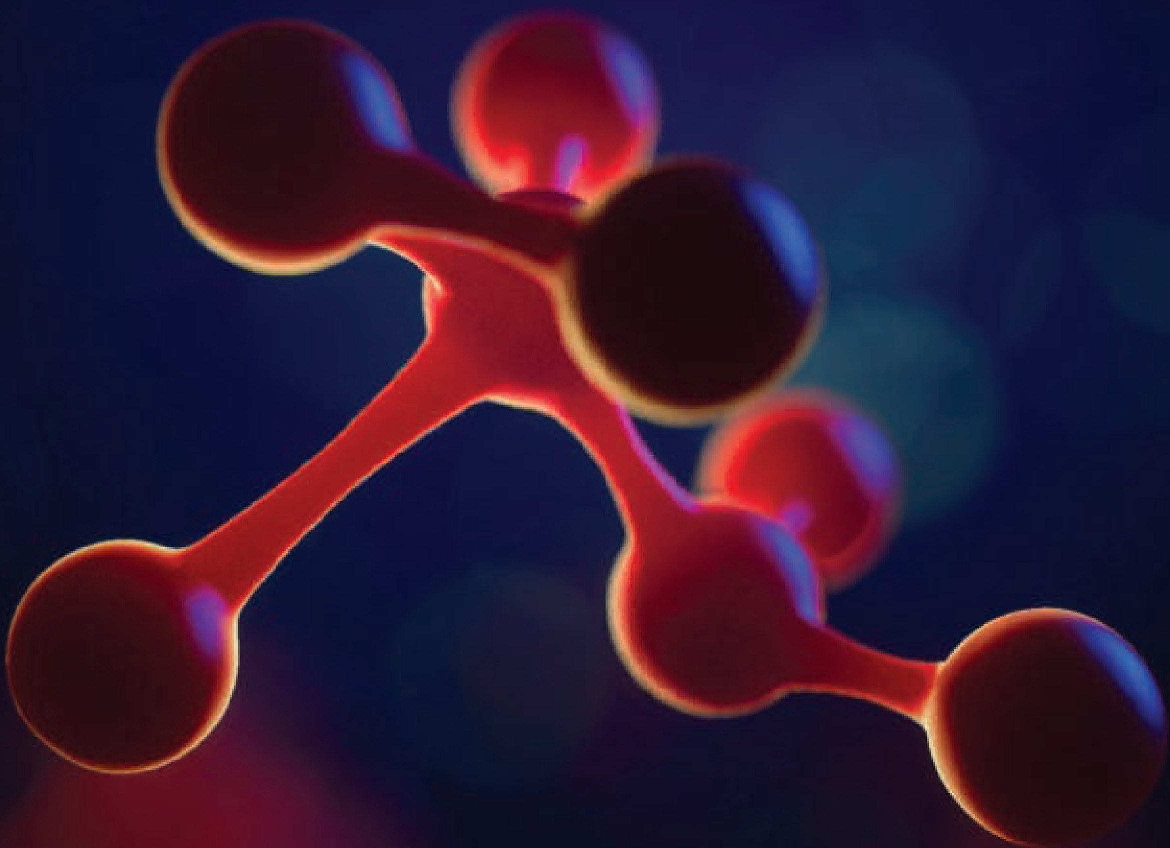
We thank K. Amarnath, R. Bundschuh, K. Jespen, C. Liu, H. Okano, and H. Zheng for helpful discussions and I. Artsimovitch, K. Fredrick, H. Garcia, L. Gerosa, I. Golding, M. Ibba, K. Kochanowski, N. Mitarai, M.-A. Moran, R. Phillips, and M. Scott for critical comments on the manuscript. RNA-seq was performed at the Institute for Genomic Medicine at UC San Diego. **Funding:** This work was supported by NSF grant MCB1818384 (to T.H.) and NIH grant R01GM109069 (to T.H.). **Author contributions:** R.B., M.M., and T.H. conceived the study, interpreted results, and wrote the manuscript. R.B. designed and performed the experiments. M.M. performed the computational and mathematical analysis of the data. Z.Z. generated the strains. I.S. developed the RNA-seq analysis pipeline. M.M., R.A., and C.L. provided and analyzed proteomic data. T.H. supervised the work and acquired funding. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The proteomic mass spectrometry files as well as data analysis files have been published previously (*10*) and are accessible through the ProteomeXchange Consortium via the PRIDE partner repository: https://www.ebi.ac.uk/pride/archive/projects/PXD014948. The *E. coli* spectral library used for DIA/SWATH data analysis is available via SWATHAtlas: http://www.peptideatlas.org/PASS/PASS01421. The genome-wide parameters of gene expression generated in this study, including transcription and translation initiation rates; mRNA degradation rates; promoter on rates; and protein, mRNA, and gene concentrations for the reference condition (glucose minimal medium) and slow glucose-limited growth, are reported in table S6. Additional analysis, including modeling and details on the calculation of the parameters, are included in the supplementary text, notes S1 to S5. Raw RNA-seq reads are uploaded to the Gene Expression Omnibus (GEO) (accession no. GSE205717). Strains generated in this study can be requested from the corresponding author. **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abk2066
Materials and Methods
Supplementary Text
Figs. S1 to S14
Tables S1 to S7
References (*55*–*110*)
MDAR Reproducibility Checklist

Submitted 28 June 2021; accepted 10 November 2022
10.1126/science.abk2066

Science
JOURNALS AAAS



Publish your research in the *Science* family of journals

The *Science* family of journals (*Science, Science Advances, Science Immunology, Science Robotics, Science Signaling*, and *Science Translational Medicine)* are among the most highly-regarded journals in the world for quality and selectivity. Our peer-reviewed journals are committed to publishing cutting-edge research, incisive scientific commentary, and insights on what's important to the scientific world at the highest standards.

Submit your research today!

Learn more at **Science.org/journals**

RESEARCH ARTICLES

COMPUTER SCIENCE

Human-level play in the game of *Diplomacy* by combining language models with strategic reasoning

Meta Fundamental AI Research Diplomacy Team (FAIR)†, Anton Bakhtin[1]‡, Noam Brown[1]*‡, Emily Dinan[1]*‡, Gabriele Farina[1], Colin Flaherty[1]‡, Daniel Fried[1,2], Andrew Goff[1], Jonathan Gray[1]‡, Hengyuan Hu[1,3]‡, Athul Paul Jacob[1,4]‡, Mojtaba Komeili[1], Karthik Konath[1], Minae Kwon[1,3], Adam Lerer[1]*‡, Mike Lewis[1]*‡, Alexander H. Miller[1]‡, Sasha Mitts[1], Adithya Renduchintala[1]‡, Stephen Roller[1], Dirk Rowe[1], Weiyan Shi[1,5]‡, Joe Spisak[1], Alexander Wei[1,6], David Wu[1]‡, Hugh Zhang[1,7]‡, Markus Zijlstra[1]

Despite much progress in training artificial intelligence (AI) systems to imitate human language, building agents that use language to communicate intentionally with humans in interactive environments remains a major challenge. We introduce Cicero, the first AI agent to achieve human-level performance in *Diplomacy*, a strategy game involving both cooperation and competition that emphasizes natural language negotiation and tactical coordination between seven players. Cicero integrates a language model with planning and reinforcement learning algorithms by inferring players' beliefs and intentions from its conversations and generating dialogue in pursuit of its plans. Across 40 games of an anonymous online *Diplomacy* league, Cicero achieved more than double the average score of the human players and ranked in the top 10% of participants who played more than one game.

A major long-term goal for the field of artificial intelligence (AI) is to build agents that can plan, coordinate, and negotiate with humans in natural language. Although much progress has been made in language models that imitate human language (*1*), effective negotiation agents must go beyond this by understanding the beliefs, goals, and intentions of their partner; planning joint actions that account for their partner's goals; and persuasively and intentionally communicating these proposals.

We present Cicero, an AI agent that achieved human-level performance in the strategy game *Diplomacy*. In *Diplomacy*, seven players conduct private natural language negotiations to coordinate their actions to both cooperate and compete with each other. By contrast, prior major successes for multi-agent AI have been in purely adversarial environments such as chess (*2*), Go (*3*), and poker (*4*), in which communication has no value. For these reasons, *Diplomacy* has served as a challenging benchmark for multi-agent learning (*5–8*).

[1]Meta AI, 1 Hacker Way, Menlo Park, CA, USA. [2]Language Technologies Institute, Carnegie Mellon University, Pittsburgh, PA, USA. [3]Department of Computer Science, Stanford University, Stanford, CA, USA. [4]Computer Science and Artificial Intelligence Laboratory, Massachusetts Insititute of Technology, Cambridge, MA, USA. [5]Department of Computer Science, Columbia University, New York, NY, USA. [6]Department of Computer Science, University of California, Berkeley, Berkeley, CA, USA. [7]EconCS Group, Harvard University, Cambridge, MA, USA.
***Corresponding author. Email: noambrown@meta.com (N.B.); edinan@meta.com (E.D.); alerer@meta.com (A.L.); mikelewis@meta.com (M.L.)** †FAIR consists of all listed authors. There are no additional authors or collaborators. ‡These authors contributed equally to this work.

Cicero couples a controllable dialogue module with a strategic reasoning engine. At each point in the game, Cicero models how the other players are likely to act on the basis of the game state and their conversations. It then plans how the players can coordinate to their mutual benefit and maps these plans into natural language messages.

We entered Cicero anonymously in 40 games of *Diplomacy* in an online league of human players between 19 August and 13 October 2022. Over the course of 72 hours of play that involved sending 5277 messages, Cicero ranked in the top 10% of participants who played more than one game.

Challenges of human-AI cooperation in *Diplomacy*

Almost all prior AI breakthroughs in games have been in two-player zero-sum (2p0s) settings, including chess (*2*), Go (*3*), heads-up poker (*9*, *10*), and *StarCraft* (*11*, *12*). In finite 2p0s games, certain reinforcement learning (RL) algorithms that learn by playing against themselves—a process known as self-play—will converge to a policy that is unbeatable in expectation in balanced games (*13*). In other words, any finite 2p0s game can be solved through self-play with sufficient compute and model capacity.

However, in games that involve cooperation, self-play without human data is no longer guaranteed to find a policy that performs well with humans, even with infinite compute and model capacity, because the self-play agent may converge to a policy that is incompatible with human norms and expectations. This effect can be clearly seen in settings that involve language, in which prior work found that self-play produced uninterpretable language despite achieving high task success for the agents (*14*, *15*). Even in dialogue-free versions of *Diplomacy*, we found that a self-play algorithm that achieved superhuman performance in 2p0s versions of the game performed poorly in games with multiple human players owing to learning a policy inconsistent with the norms and expectations of potential human allies (*16*, *17*). Thus, a major challenge in *Diplomacy* is to develop a way to harness the potential benefits of self-play in a way that leads to human-compatible language and behavior.

The challenge of maintaining human-interpretable communication is particularly acute in *Diplomacy*, in which our agent sent and received an average of 292 messages per game (fig. S8). Messages in the game often involve coordinating precise plans, and any miscommunication can result in their failure. Each message an agent sends must be grounded in (be contextually appropriate and consistent with) lengthy dialogue histories, game states—including proposed hypothetical states—and goals. If messages are inaccurately grounded, humans may ask the agent to explain its errors (a challenging task that may lead to further mistakes) or choose to cooperate with others instead. Further, repeated messaging creates feedback loops, in which the language model imitates the style of its own previous messages—for example, sending a short or incoherent message will increase the likelihood of such messages in the future (*18*). Past work on strategic dialogue systems has avoided these issues by focusing on simpler settings (*14*, *19–21*), which involve only a single human partner, shorter dialogue histories, and simpler strategies.

Last, *Diplomacy* is a particularly challenging domain because success requires building trust with others in an environment that encourages players to not trust anyone. Each turn's actions occur simultaneously after nonbinding, private negotiations. To succeed, an agent must account for the risk that players may not stay true to their word, or that other players may themselves doubt the honesty of the agent. For this reason, an ability to reason about the beliefs, goals, and intentions of others and an ability to persuade and build relationships through dialogue are powerful skills in *Diplomacy*.

The game of *Diplomacy*

Diplomacy is a board game in which seven players compete to control supply centers (SCs) on a map, by moving their units into them. A player wins by controlling a majority of SCs. The game may also end when all remaining players agree to a draw, or a turn limit is reached, in which case scores are determined by the number of SCs each player controls.

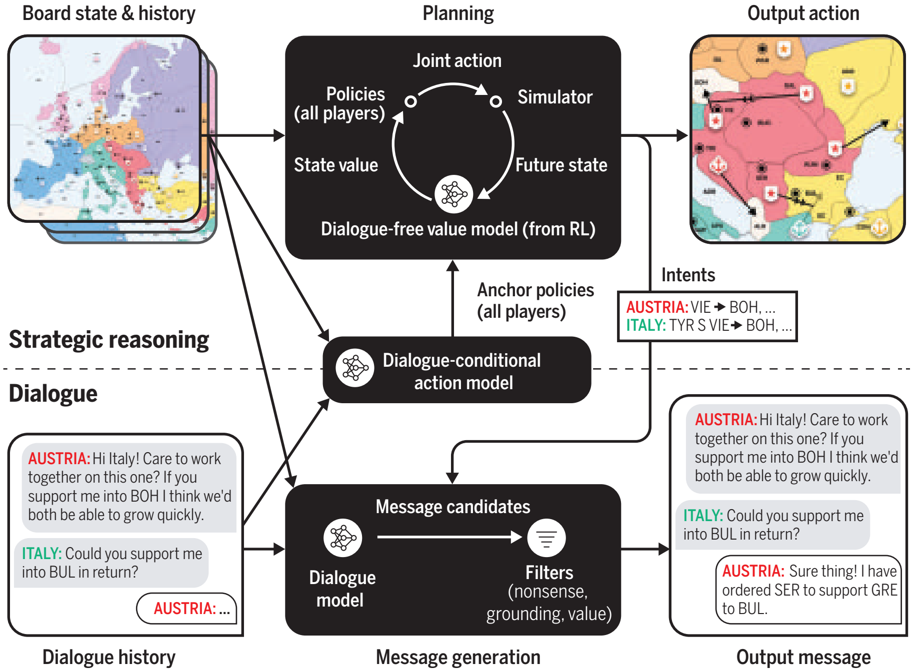


Fig. 1. Architecture of Cicero. Cicero predicts likely human actions for each player according to the board state and dialogue, using that as the starting point for a planning algorithm using RL-trained models. The output of planning is an action for the agent as well as beliefs about other players' actions, which are used to select intents for a dialogue model to condition on. Generated message candidates undergo several filtering steps before a final message is sent.

Each turn, all players engage in private pairwise free-form dialogue with the others during a negotiation period, and then all players simultaneously choose an action comprising one order per unit they control. A unit may support other units, including those of another player, which forms the basis for much of the negotiation in *Diplomacy*. A detailed description of the rules is provided in the supplementary materials (SM), materials and methods, section C.

Overview of Cicero

At a high level, Cicero combines a dialogue module with a strategic reasoning module, along with a filtering process to reject low-quality messages. A diagram of Cicero is provided in Fig. 1.

Dialogue

Cicero generates dialogue using a pretrained language model that was further trained on dialogue data from human games of *Diplomacy*. Crucially, in addition to being grounded in both the dialogue history and game state, the dialogue model was trained to be controllable through intents, which we here define to be a set of planned actions for the agent and its speaking partner. This was accomplished by automatically augmenting the human data with inferred intents and using this information as further conditioning during training. For example, intents showing the agent moving into the territory Bulgaria ("BUL") with support from its speaking partner might yield a message such as "Could you support me into BUL in return?" Grounding in intents relieved the dialogue model of most of the responsibility for learning which actions were legal and strategically beneficial. In particular, this control provided an interface between the dialogue generation and strategic reasoning.

Strategic reasoning

Cicero uses a strategic reasoning module to intelligently select intents and actions. This module runs a planning algorithm that predicts the policies of all other players on the basis of the game state and dialogue so far, accounting for both the strength of different actions and their likelihood in human games, and chooses an optimal action for Cicero that is based on those predictions. Planning relies on a value and policy function trained through self-play RL that penalized the agent for deviating too far from human behavior, to maintain a human-compatible policy. During each negotiation period, intents are recomputed every time Cicero sends or receives a message. At the end of each turn, Cicero plays its most recently computed intent.

Message filtering

Cicero passes each generated message through several filters designed to limit messages that are nonsensical, inconsistent with intents, or strategically poor.

Methods

Data

We obtained a dataset of 125,261 games of *Diplomacy* played online at webDiplomacy.net. Of these, 40,408 games contained dialogue, with a total of 12,901,662 messages exchanged between players. Player accounts were de-identified, and automated redaction of personally identifiable information (PII) was performed by webDiplomacy. We refer to this dataset hereafter as WebDiplomacy.

Intent-controlled dialogue

Cicero generates messages through a neural generative Diplomacy dialogue model that was trained to be controllable through a set of intents.

Imitation dialogue model

We took R2C2 (*22*) as our base model—a 2.7 billion–parameter Transformer-based (*23*) encoder-decoder model pretrained on text from the internet by using a BART denoising objective (*24*). The base pretrained model was then further trained on WebDiplomacy (Methods, Data) through standard maximum likelihood estimation. Specifically, with a dataset $\mathcal{D} = \{[\mathbf{x}^{(i)}, \mathbf{y}^{(i)}]\}$, the model was trained to predict a dialogue message $\mathbf{y}^{(i)}$ from player $\mathcal{A}$ to player $\mathcal{B}$ at time t, given all of the following represented as text $\mathbf{x}^{(i)}$: dialogue history (all messages exchanged between player $\mathcal{A}$ and the six other players up to time t); game state and action history (current game state and recent action history); player rating (rating for $\mathcal{A}$ corresponding to Elo rating computed from games in WebDiplomacy); game and message metadata (additional info about game settings and the current message, such as time since the last message, and current turn). Additionally, the model conditions on intents (a set of proposed actions for players $\mathcal{A}$ and $\mathcal{B}$ for the current turn and future turns, representing the intent for message $\mathbf{y}^{(i)}$). Further details on the training data, training procedure, relevant hyperparameters, sampling procedures, and other inference-time methods are provided in the SM, section D.1.

During play, we used additional modules governing when to speak and to whom, which are described in the SM, section D.4.

Controllable dialogue model through intents

Standard language modeling approaches would train our dialogue model only to imitate the messages from our dataset but not to outperform them. To go beyond imitation learning, we made the dialogue model controllable by generating messages conditioned on a plan specified by the strategic reasoning module

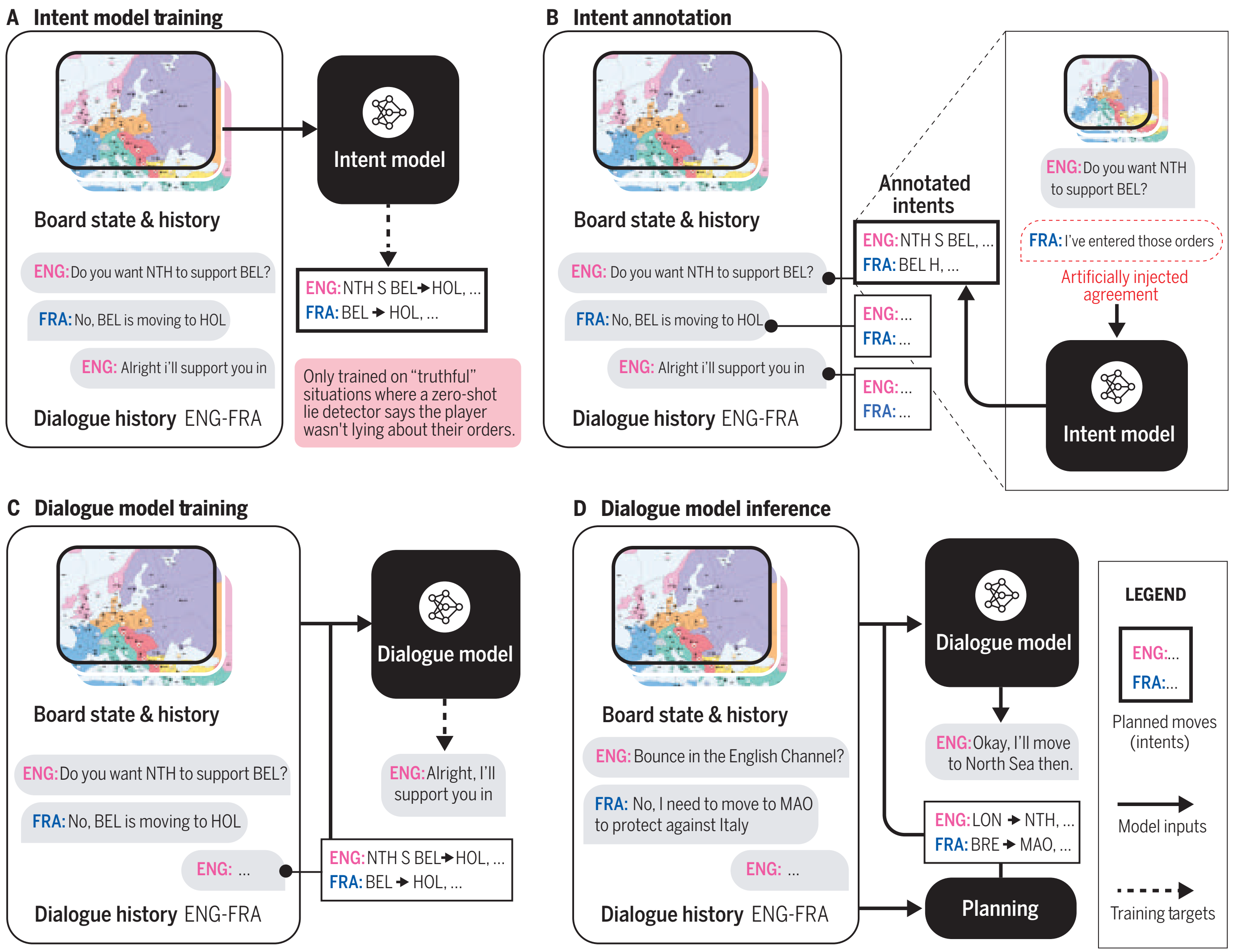


Fig. 2. Illustration of the training and inference process for intent-controlled dialogue. Actions are specified as strings of orders for units; for example, "NTH S BEL - HOL" means that North Sea will support Belgium to Holland. (**A**) An "intent model" was trained to predict actions for a pair of players on the basis of their dialogue. Training data was restricted to a subset in which dialogue is deemed "truthful" (SM, section D.2.3). (**B**) Each message in the dialogue training dataset was annotated with the output of the intent model on the dialogue up to that point, with an agreement message injected at the end. (**C**) The dialogue model was trained to predict each dataset message given the annotated intent for the target message. (**D**) During play, intents were supplied by the planning module instead.

(intents), resulting in higher-quality messages. More specifically, a message is defined to have intent **z** if **z** is the most likely set of actions that the sender and recipient will take—for both the current turn and several future turns—if no further dialogue occurs after the message is received. To establish this control, we developed techniques to automatically annotate every message in the training set with a set of actions corresponding to the message content. During training, the dialogue model learned the distribution $p_\theta\left[\mathbf{y}^{(i)}\middle|\mathbf{x}^{(i)},\mathbf{z}^{(i)}\right]$, where $\mathbf{z}^{(i)}$ represents the intent for datapoint $[\mathbf{x}^{(i)},\mathbf{y}^{(i)}]$; as a result, at inference, time **z** provides a point of control over generation (*25*). We later describe the training and inference process, which is also illustrated in the pipeline in Fig. 2. The effect of the intents on the generated dialogue is demonstrated in Fig. 3; conditioning on different planned actions results in different messages.

We considered other notions of intent during development, such as controlling messages to focus on specific subsets of actions, third-party actions, or to have a particular tone. Richer intents are harder to annotate on human messages, are harder to select with the planning module, and create greater risk of taking the language model out of distribution.

Annotating training messages with intents

When annotating messages in the training data with corresponding intents, our goal was for the proposed actions $\mathbf{z}^{(i)}$ to closely reflect the content of a message $\mathbf{y}^{(i)}$ so that at training time, the model learned to exploit the information in $\mathbf{z}^{(i)}$.

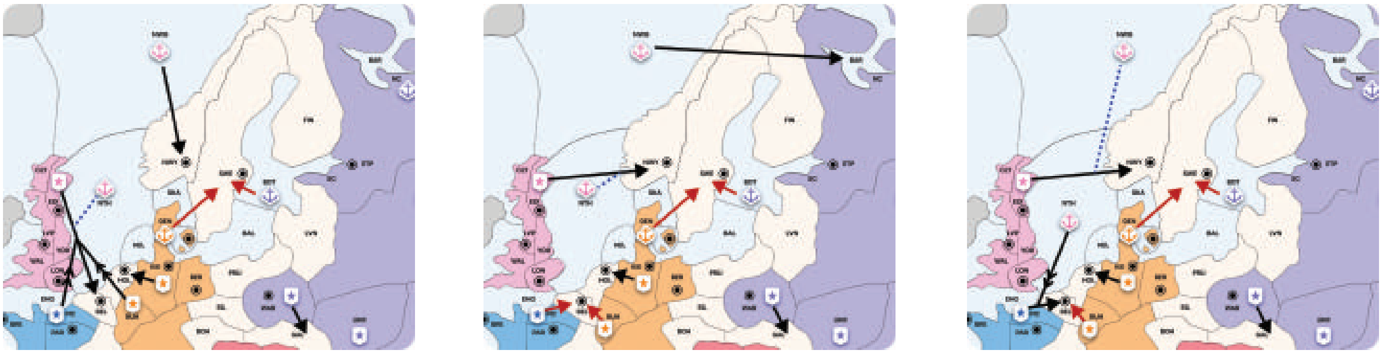
Naïvely, we could have used the actual actions played by the sender and recipient at the end of each turn in the span of the intent. However, these actions may not reflect the content of a message if (i) a message is not honest or (ii) subsequent messages change the sender's plans. To resolve (i), we predicted the most likely action according to a dialogue-conditional action prediction model trained on a "truthful" subset of the dataset, in which we predicted that a player's dialogue was not deceptive to others (SM, section D.2). This is showcased in Fig. 2A; we refer to this model as the intent model. To resolve (ii), we restricted the dialogue history that this intent model saw up to the message in question, which signaled to the model to predict actions as though the dialogue had ended at that point in time. We

Depicted intents:

England convoys an army to Belgium with the support of France and Germany while taking Norway in a manner friendly to Russia.

England convoys an army to Norway and also aggressively moves to the Barents Sea against Russia.

England convoys an army to Norway and supports France to gain Belgium, preventing Germany from taking Belgium.



Example generations from the dialogue model with these actions as the conditioned intent:

ENG→ FRA Mind supporting Edi - Bel?

ENG→ GER Do you wanna support my convoy to Bel? With Italy going aggressive France will fall quickly and we can make gains off of both Russia and France.

ENG→ RUS How are you thinking Germany is gonna open? I may have a shot at Belgium, but I'd need your help into Den next year.

ENG→ FRA What's the plan for Bel?

ENG→ GER I will be convoying my army to Norway and moving to Barrents Sea.

ENG→ RUS Hey, I moved to Edi to keep my options open. Do you have a problem with me putting an army in Nwy?

ENG→ FRA Let me know if you want me to support you to Bel, otherwise I'll probably poke Hol.

ENG→ GER Looks like you'll get three builds unless France bounces you! Are you gonna bounce Russia or not?

ENG→ RUS Hey, I moved to Edi to keep my options open. Do you have a problem with me putting an army in Nwy?

Fig. 3. The effect of intents on Cicero's dialogue. Pictured are three different possible intents in the same game situation. In each case, we show a message generated by Cicero (England; pink) to France (blue), Germany (orange) and Russia (purple) conditioned on these intents. Each intent leads to quite different messages, which are consistent with the intended actions.

additionally added messages to the dialogue history that suggested a conclusive agreement between the two parties (Fig. 2B). As a result, we obtained a high degree of correspondence between the action annotated as the intent of a message and the content, achieving a score of 97% on a small test set designed to measure this correspondence (compared with 77% for a simpler baseline) (table S2). Then, the dialogue model could be trained in the manner described in the above section Imitation dialogue model and in Fig. 2C (SM, section D.2).

Selecting intents during play

During play, Cicero used the strategic reasoning module to select intent actions for the current turn (Fig. 2D), whereas intent actions for future turns were generated by means of a human-imitation model.

Agent intent action for current turn

Cicero conditioned its dialogue on the action that it intends to play for the current turn. This choice maximizes Cicero's honesty and its ability to coordinate but risks leaking information that the recipient could use to exploit it (for example, telling them which of their territories Cicero plans to attack) and sometimes led to out-of-distribution intents when the intended action was hostile, because in adversarial situations, humans may rarely communicate their intent honestly. We describe approaches for mitigating these risks in the section Message filtering.

Recipient intent action for current turn

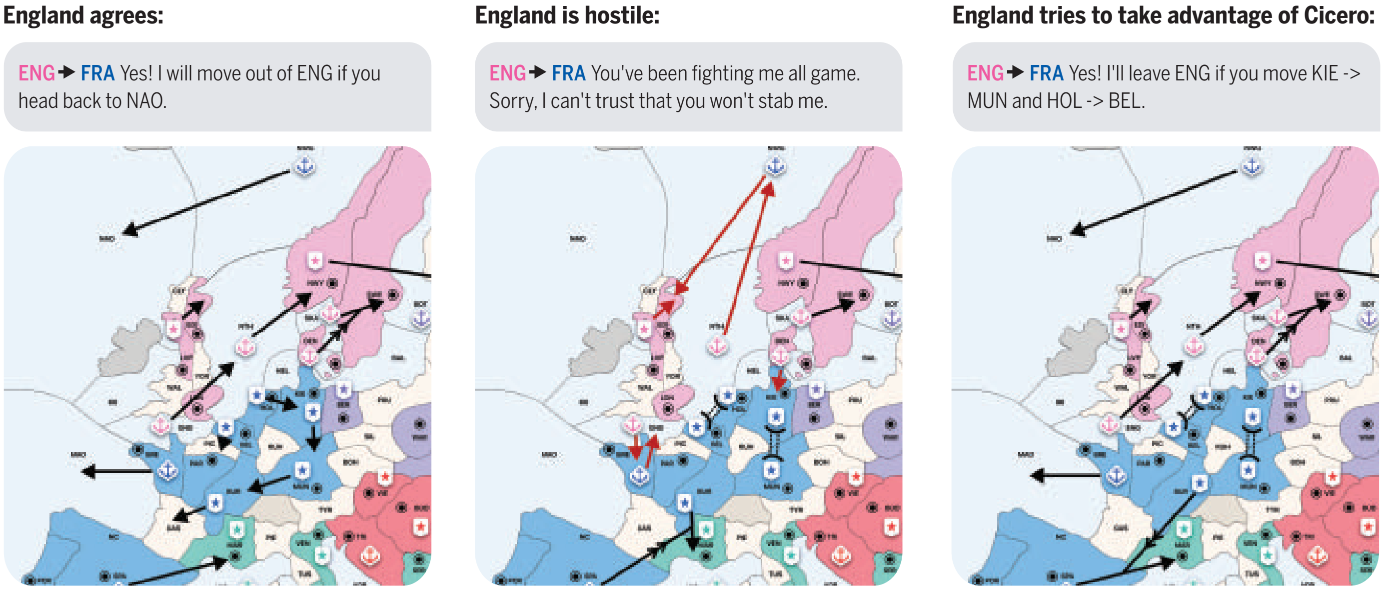
Cicero considered the subset of recipient actions with high likelihood under its beliefs about their policy. High likelihood requires that either an action is deemed beneficial for the recipient and/or that they are believed to be likely to play it given the dialogue. Among this restricted set, Cicero selected the recipient action with the highest expected value for itself (SM, section D.2.4).

Dialogue modeling results

We compared the performance of our dialogue model with a baseline without intent grounding and one without intent or game-state grounding (a "language model"). We report

	DIALOGUE QUALITY RATINGS (%)			
	Consistent with state	Consistent with plan	High quality	Perplexity
Language model	61.90	76.19	20.64	8.02
+ game state grounding	84.13	83.33	29.37	7.94
+ intent grounding (CICERO)	87.30	92.86	37.30	7.70

Fig. 4. Controllable dialogue modeling results. We report dialogue quality ratings and perplexity on the validation set for the Cicero dialogue model and compare them with a baseline without intent grounding and a baseline without either intent or game-state grounding ("Language model"). Dialogue quality ratings were calculated according to expert annotation of generated messages in 126 situations; we report the percent of messages (before filtering) labeled as consistent with the game state, as consistent with the plan for the next actions, and as particularly high quality. Lower perplexity corresponds to more probability mass on the ground-truth human messages.



Cicero predicts England will retreat from ENG to NTH 85% of the time, backs off its own fleet to NAO as agreed, and begins to move armies away from the coast.

Cicero does not back off its fleet but rather attacks EDI with it, and leaves its armies at the coast to defend against an attack from England, predicting that England will attack about 90% of the time.

Strategic planning rejects the possibility of vacating KIE and HOL, because it would make Cicero too vulnerable. Cicero backs off its fleet to NAO but keeps armies at the coast to defend.

Fig. 5. The effect of dialogue on Cicero's strategic planning and intents. Cicero (France; blue) and England (pink) are entangled in a fight, but it would be beneficial for both players if they could disengage. Cicero has just messaged England "Do you want to call this fight off? I can let you focus on Russia and I can focus on Italy." Pictured are three ways that England might reply and how Cicero adapts to each. (**Left** and **middle**) Because Cicero's planning anchors around a dialogue-conditional policy model, its predictions for other players and accordingly its own plans are flexible and responsive to negotiation with other players. (**Right**) Yet Cicero also avoids blindly trusting what other players propose by rejecting plans that have low predicted value and run counter to its own interests.

both perplexity on the validation set and dialogue quality rating scores, which were calculated on the basis of expert annotation of messages generated in 126 *Diplomacy* game situations. Experts were asked to label whether a message was (i) consistent with the game state, (ii) consistent with the agent's plan, and (iii) notably high quality, compared with that of an average human. Results are shown in Fig. 4, and more details regarding this evaluation are provided in the SM, section D.2.3. Our model outperformed the baselines on all metrics. The improvement in validation perplexity demonstrated that the model can use additional grounding information to better predict human messages. Expert annotations showed that the grounding information provided by the intents and game state led to higher-quality messages that were highly consistent with the agent's intended action.

Strategic reasoning

To generate the intents for dialogue and to choose the final actions to play each turn, Cicero ran a strategic reasoning module that predicts other players' policies (a probability distribution over actions) for the current turn according to the state of the board and the shared dialogue and then chose a policy for itself for the current turn that responded optimally to the other players' predicted policies.

Doing this with human players requires predicting how humans will play. A popular approach in cooperative games is to model the other players' policies through supervised learning on human data, which is commonly referred to as behavioral cloning (BC). However, pure BC is brittle, especially because a supervised model may learn spurious correlations between dialogue and actions (fig. S6).

To address this problem, Cicero used variants of piKL (*26*) to model the policies of players. piKL is an iterative algorithm that predicts policies by assuming each player i seeks to both maximize the expected value of their policy π_i and minimize the Kullback-Leibler (KL) divergence between π_i and the BC policy, which we call the anchor policy τ_i. An anchor strength parameter $\lambda \in [0, \infty)$ trades off between these competing objectives.

piKL: KL-regularized planning

piKL is an iterative algorithm that predicts player policies. A complete description of the algorithm can be found in the SM, section E.1. piKL treats each turn in *Diplomacy* as its own subgame in which each player i simultaneously chooses an action a_i that results in joint action $a = (a_1, ..., a_n)$, and then each player i receives a reward $u_i(a)$ determined by a value function u_i. We discuss the training of this value function later below.

piKL assumes player i seeks a policy π_i that maximizes the modified utility function

$$U_i(\pi_i, \pi_{-i}) = u_i(\pi_i, \pi_{-i}) - \lambda D_{KL}(\pi_i \| \tau_i) \quad (1)$$

where π_{-i} represents the policies of all players other than i, and $u_i(\pi_i, \pi_{-i})$ is the expected value of π_i given that other players play π_{-i}. Specifically, let $Q_i^{t-1}(a_i) = u_i\left(a_i, \pi_{-i}^{t-1}\right)$ and let

$$\pi_i^{\Delta t}(a_i) \propto \tau_i(a_i)\exp\left[\frac{Q_i^{t-1}(a_i)}{\lambda}\right] \quad (2)$$

On each iteration t, piKL updates its prediction of the players' joint policies to be

$$\pi^t = \left(\frac{t-1}{t}\right)\pi^{t-1} + \left(\frac{1}{t}\right)\pi^{\Delta t} \quad (3)$$

piKL provably converges to an equilibrium in the modified utility space (*26*). When the anchor strength λ is set to a large value, piKL predicts that player i's policy will be close to the anchor policy τ_i. When λ is small, piKL predicts that player i's policy will have high expected value and may deviate substantially from τ_i.

A generalization of piKL referred to as Distributional Lambda piKL (DiL-piKL) replaces the single λ parameter in piKL with a probability distribution over λ values (SM, section E.1.3). On each iteration, each player samples a λ value from their distribution. In practice, we found this led to better performance (*17*).

Dialogue-conditional planning

Because dialogue influences the BC policy (the anchor policy τ_i), piKL provides a mechanism for dialogue to influence policy predictions. Different possible messages between Cicero and another player may produce different anchor policies (Fig. 5), which ultimately gives different final predictions about what that player will do.

Other players may of course be deceptive about their plans. Cicero does not explicitly predict whether a message is deceptive or not but rather relies on piKL to directly predict the policies of other players on the basis of both the BC policy (which conditions on the message) and on whether deviating from the BC policy would benefit that player.

Because dialogue in *Diplomacy* occurs privately between pairs of players, Cicero must reason about what information players have access to when making predictions. For example, if Cicero is coordinating an attack with an ally against an adversary, Cicero's prediction of the adversary's policy must account for the adversary not being aware of the intended coordination. Cicero accomplished this by predicting by means of pairwise piKL what every other player's policy will be.

Specifically, during strategic planning, for each player j, Cicero computed an anchor policy for both itself and player j on the basis of their shared conversation, the board state, and the recent action history. Cicero then ran DiL-piKL for the two players to predict player j's policy. On each iteration, Cicero assumed that the remaining five players would play according to a policy computed by means of RL, conditional on the policies of Cicero and player j. This process gave an independent prediction of each player's policy.

Next, Cicero accounted for the players' policies not being independent owing to their ability to correlate their actions through private dialogue that Cicero did not observe. Cicero accomplished this by constructing an approximate joint policy for all other players through self-normalized importance sampling: We sampled N = 1000 joint actions a from the independent piKL policies of the other players and reweighted them by the likelihood ratio of a under the correlated and independent RL policies, respectively.

Last, Cicero chose the action a_i that best responds to the predicted joint policy π_{-i} of the other players, while still being as consistent as possible with its dialogue. Specifically, Cicero chose the action $\text{argmax}_{a_i} u_i(a_i, \pi_{-i}) + \lambda \log \tau_i(a_i)$, where u_i is the RL value function, $\tau_i(a_i)$ is the probability of the action under the dialogue-conditional imitation policy, and $\lambda = 3 \times 10^{-3}$. Cicero used a smaller λ for regularizing its best response than for its computation of other players' policies; thus, the dialogue more strongly informed Cicero's expectations of how other players would coordinate while still allowing Cicero more leeway to deviate when the action that it predicted humans would most likely choose in its situation was suboptimal.

Self-play RL for improved value estimation

Applying piKL requires a state value function. Self-play provides an avenue for training such a value function but risks becoming incompatible with human play (*16*, *17*). To address this, we used piKL during self-play to keep the policies human-compatible.

One challenge in doing self-play in *Diplomacy* is that players may adapt their actions substantially on the basis of dialogue with other players, including coordinating joint actions. Explicitly simulating conversations would be extremely expensive in RL. However, a key insight is that a joint, shared BC policy trained on the joint action distribution of the human data already implicitly captures the effects of dialogue on the action distribution of human players by modeling that action distribution directly.

We therefore developed Correlated and Shared (CoShar) piKL, which allowed for regularization toward a joint, correlated anchor policy τ shared by all players rather than toward per-player policies. In this way, we relied on the joint anchor policy to capture the correlation between all players' policies. Specifically, CoShar piKL differs from standard piKL in that the probability of joint action a = $(a_1, ..., a_n)$ in policy $\pi^{\Delta t}$ becomes

$$\pi^{\Delta t}(a) \propto \tau(a) \exp\left[\sum_{i \le n} \frac{Q_i^{t-1}(a_i)}{\lambda}\right] \tag{4}$$

We found that CoShar piKL retained much of the correlation present in the joint anchor policy τ while also modeling strong human play better than imitation alone.

Our resulting self-play algorithm operated similarly to AlphaZero (*27*) and ReBeL (*28*) by applying planning "in the loop" as the improvement operator for RL. In our case, planning was by use of an approximated version of CoShar piKL. We generated self-play trajectories in which on each turn we computed the CoShar piKL policy using a learned state-value model. We regressed the joint policy model toward that policy and regressed the value model toward the expected values of all players under that policy. We then sampled a joint action from that policy to generate the next state in the trajectory. The anchor policy was fixed throughout training to anchor the RL near human play (SM, section E.4).

Message filtering

Prior work has shown that neural language models suffer from contradictions and inconsistency as well as a tendency to "hallucinate," or generate factually incorrect information (*29*). In the complex domain of *Diplomacy*, dialogue models exhibit both these problems and other more subtle mistakes, such as deviations from the intents used to control the message or blunders in the strategic content of the message. We approached this problem by filtering generated messages using a series of classifiers and checks to detect common issues. We outline

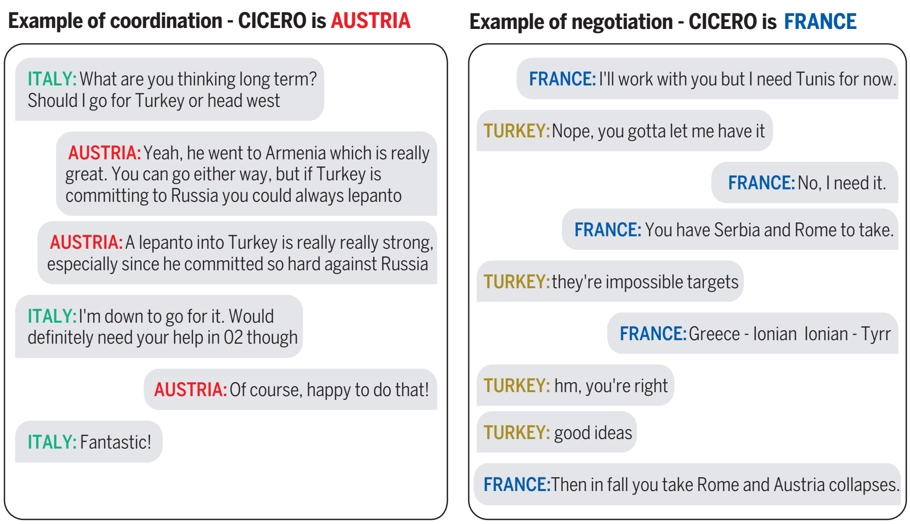


Fig. 6. Successful dialogue examples. Examples of Cicero (**left**) coordinating and (**right**) negotiating with authors of this paper in test games.

several of these filters here, with additional details in the SM, section D.3.

Discriminating between human text and counterfactuals

Much work has used adversarial or counterfactual examples to improve the robustness of natural language systems (*30*, *31*). Following this approach, we generated many kinds of counterfactual messages that contained mistakes that language models are prone to, including heuristically corrupted text as well as model-generated negatives. We trained a suite of 16 classifiers to discriminate between the ground-truth human message and different kinds of counterfactual messages (sometimes varying the random seed or context information available) and used these classifiers in an ensemble to filter messages. This approach risked overly filtering complex messages that contain precise plans and accepting bland messages, such as “ok,” which are unlikely to contain mistakes. However, we found that carefully designing our ensemble allowed us to filter most nonsensical messages with minimal impact on message complexity: On a small evaluation set with 362 expert-annotated examples, we found that we could detect 83% of nonsense messages, without substantial impact to message diversity as measured by the proxy of message length and the number of references to *Diplomacy*-specific entities (SM, section D.3.1).

Intent correspondence

As noted previously, controlling dialogue generation through intents has the twofold benefit of improving the strategic value of a message and reducing discussion of impossible moves or other hallucinations. However, this control is imperfect, and the dialogue model may generate messages that contradict the intents it conditions on. To address this, we filtered messages that would reduce the likelihood of the actions in the intent. Evaluating this method on a small test set of 1013 expert-annotated messages, we achieved a recall of 65%, filtering 24% of all messages (SM, section D.3.2).

Value-based filtering

Conditioning on intents can lead to “information leakage,” in which the agent reveals compromising information about its plan to an adversary (section Selecting intents during play). To mitigate this, we developed a method to score potential messages by their estimated value impact. We computed the piKL policies for all agents after each candidate message and filtered those that led to a lower expected value (EV) for Cicero playing its intended action. Expert evaluation on a set of 127 dialogue scenarios demonstrated that accepted messages were preferred over filtered messages 62% of the time ($P < 0.05$) (SM, section D.3.3).

Other filters

We additionally deployed other filters—for example, to detect toxic language (SM, section D.3.4)—and heuristics to curb bad behaviors, including repetition and off-topic messages (SM, section D.3.5).

Cicero in anonymous human play

Cicero participated anonymously in 40 games of *Diplomacy* in a “blitz” league on webDiplomacy.net from 19 August to 13 October 2022. This league played with 5-min negotiation turns; these time controls allowed games to be completed within 2 hours. Cicero ranked in the top 10% of participants who played more than one game and second out of 19 participants in the league that played five or more games. Across all 40 games, Cicero’s mean score was 25.8%, which was more than double the average score of 12.4% of its 82 opponents. As part of the league, Cicero participated in an eight-game tournament that involved 21 participants, six of whom played at least five games. Participants could play a maximum of six games, with their rank determined by the average of their best three games. Cicero placed first in this tournament.

During games, players were not able to see the usernames of other players. Although webDiplomacy notifies users that the website has participated in AI research and that certain game modes allow users to play with AI agents, we evaluated Cicero in games with humans in which the participants were not explicitly informed that they were playing with an AI agent for that particular game. Cicero’s participation as an AI was revealed to all players at the conclusion of the research (SM, section A.4).

Discussion

Cicero successfully combined strategic reasoning and dialogue to cooperate and negotiate with humans on a complex task, achieving strong human-level performance in the game of *Diplomacy*. Furthermore, Cicero passed as a human player for 40 games of *Diplomacy* with 82 distinct players, and no in-game messages indicated that players believed that they were playing with an AI agent. One player mentioned in post-game chat a suspicion that one of Cicero’s accounts might be a bot, but this did not lead to Cicero being detected as an AI agent by other players in the league.

Two examples of coordination and negotiation are shown in Fig. 6. In the coordination example, we observed Cicero building an alliance through discussion of a longer-term strategy. In the negotiation example, Cicero successfully changed the other player’s mind by proposing mutually beneficial moves. Despite dishonesty being commonplace in *Diplomacy*, we were able to achieve human-level performance by controlling the agent’s dialogue through the strategic reasoning module to be largely honest and helpful to its speaking partners.

Although Cicero is shown to be effective at cooperating with humans, it occasionally sent messages that contained grounding errors, contradicted its plans, or were otherwise strategically subpar. Although we reduced errors with a suite of filters, *Diplomacy* poses an interesting benchmark for studying this problem. We suspect that these mistakes did not raise further suspicions that Cicero was an AI agent because of the time pressure imposed by the game, as well as because humans occasionally make similar mistakes. As such, formats of *Diplomacy* with longer negotiation periods could provide an even further challenge for future work because players typically engage in more detailed and complex negotiation in these formats.

From a strategic perspective, Cicero reasoned about dialogue purely in terms of players’ actions for the current turn. It did not model how its dialogue might affect the relationship with other players over the long-term course of a game. Considering this might allow it to deploy dialogue more strategically. Furthermore, the expressive power of our intent representation limited Cicero’s ability to control richer affordances of dialogue such as strategically revealing information, asking questions, or providing explanations for its actions. There remain many open problems for intentional use of dialogue, and *Diplomacy* provides a rich testbed to explore these connections between strategy and communication, with the goal of improving coordination between humans and agents.

Ethical considerations

We discuss ethical considerations for this research further in the SM, including privacy considerations for data usage (SM, section A.1), potential harms resulting from toxic or biased language generation (SM, section A.2), avenues for misuse of goal-oriented dialogue technology (SM, section A.3), and AI agent disclosure to human players (SM, section A.4).

REFERENCES AND NOTES

1. T. Brown *et al.*, *Adv. Neural Inf. Process. Syst.* **33**, 1877 (2020).
2. M. Campbell, A. J. Hoane Jr., F. Hsu, *Artif. Intell.* **134**, 57–83 (2002).
3. D. Silver *et al.*, *Nature* **529**, 484–489 (2016).
4. N. Brown, T. Sandholm, *Science* **365**, 885–890 (2019).
5. S. Kraus, D. Lehmann, Diplomat, an agent in a multi agent environment: An overview, in *IEEE International Performance Computing and Communications Conference* (IEEE Computer Society, 1988), pp. 434–435.
6. D. d. Jonge *et al.*, The challenge of negotiation in the game of *Diplomacy*, in *International Conference on Agreement Technologies* (Springer, 2018), pp. 100–114.
7. P. Paquette *et al.*, *Adv. Neural Inf. Process. Syst.* **32**, 4474–4485 (2019).
8. A. Dafoe *et al.*, *Nature* **593**, 33–36 (2021).
9. M. Moravčík *et al.*, *Science* **356**, 508–513 (2017).
10. N. Brown, T. Sandholm, *Science* **359**, 418–424 (2018).
11. O. Vinyals *et al.*, *Nature* **575**, 350–354 (2019).

12. Dota 2 (*32*) is two-team zero-sum but with unlimited information sharing between teammates, which makes the game equivalent to 2p0s. Prior work found that self-play from scratch was sufficient for achieving superhuman performance in multiplayer poker (*4*), but this may be due to poker offering few opportunities for players to cooperate.
13. J. v. Neumann, *Math. Ann.* **100**, 295 (1928).
14. M. Lewis, D. Yarats, Y. Dauphin, D. Parikh, D. Batra, Deal or no deal? End-to-end learning of negotiation dialogues, in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing* (Association for Computational Linguistics, Copenhagen, Denmark, 2017), pp. 2443–2453.
15. A. P. Jacob, M. Lewis, J. Andreas, Multitasking inhibits semantic drift, in *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (2021), pp. 5351–5366.
16. A. Bakhtin, D. Wu, A. Lerer, N. Brown, *Adv. Neural Inf. Process. Syst.* **32**, 34 (2021).
17. A. Bakhtin *et al.*, Mastering the game of no-press *Diplomacy* via human-regularized reinforcement learning and planning. arXiv:2210.05492 [cs.GT] (2022).
18. A. Holtzman, J. Buys, L. Du, M. Forbes, Y. Choi, The curious case of neural text degeneration, *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020* (OpenReview.net, 2020).
19. S. Keizer *et al.*, Evaluating persuasion strategies and deep reinforcement learning methods for negotiation dialogue agents, in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers* (Association for Computational Linguistics, 2017), pp. 480–484.
20. T. Hiraoka, G. Neubig, S. Sakti, T. Toda, S. Nakamura, Construction and analysis of a persuasive dialogue corpus, in *Situated Dialog in Speech-Based Human-Computer Interaction* (Springer, 2016), pp. 125–138.
21. X. Wang *et al.*, Persuasion for good: Towards a personalized persuasive dialogue system for social good, in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (Association for Computational Linguistics, Florence, Italy, 2019), pp. 5635–5649.
22. K. Shuster *et al.*, Language models that seek for knowledge: Modular search and generation for dialogue and prompt completion. arXiv:2203.13224 (2022).
23. A. Vaswani *et al.*, Attention is all you need, in *Advances in Neural Information Processing Systems*, I. Guyon, *et al.*, eds. (Curran Associates, Inc., 2017), vol. 30.
24. M. Lewis *et al.*, BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension, in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5–10,2020*, D. Jurafsky, J. Chai, N. Schluter, J. R. Tetreault, Eds. (Association for Computational Linguistics, 2020), pp. 7871–7880.
25. N. S. Keskar, B. McCann, L. R. Varshney, C. Xiong, R. Socher, CTRL: A conditional transformer language model for controllable generation. arXiv:1909.05858 [cs.CL] (2019).
26. A. P. Jacob *et al.*, Modeling strong and human-like gameplay with KL-regularized search, in *International Conference on Machine Learning* (PMLR, 2022), pp. 9695–9728.
27. D. Silver *et al.*, *Science* **362**, 1140–1144 (2018).
28. N. Brown, A. Bakhtin, A. Lerer, Q. Gong, *Adv. Neural Inf. Process. Syst.* **33**, 17057 (2020).
29. Z. Ji *et al.*, Survey of hallucination in natural language generation. arXiv:2202.03629 [cs.CL] (2022).
30. P. Gupta, Y. Tsvetkov, J. P. Bigham, Synthesizing adversarial negative responses for robust response ranking and evaluation, in *Findings of the Association for Computational Linguistics: ACL/IJCNLP 2021, Online Event, August 1-6,2021*, C. Zong, F. Xia, W. Li, R. Navigli, Eds. (Association for Computational Linguistics, 2021), vol. ACL/IJCNLP 2021 of *Findings of ACL*, pp. 3867–3883.
31. M. Alzantot *et al.*, Generating natural language adversarial examples, in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31–November 4,2018*, E. Riloff, D. Chiang, J. Hockenmaier, J. Tsujii, eds. (Association for Computational Linguistics, 2018), pp. 2890–2896.
32. C. Berner *et al.*, Dota 2 with large scale deep reinforcement learning. arXiv:1912.06680 (2019).
33. FAIR *et al.*, Supplementary data for "Human-level play in the game of *Diplomacy* by combining language models with strategic reasoning". *Zenodo* (2022).
34. FAIR *et al.*, Code for "Human-level play in the game of *Diplomacy* by combining language models with strategic reasoning". *GitHub* (2022); https://github.com/facebookresearch/diplomacy_cicero.

ACKNOWLEDGMENTS

We thank K. Kuliukas for providing access to the WebDiplomacy data and for supporting this research; J. Andreas, N. Goyal, P. Paquette, K. Shuster, and S. Zhang for helpful support and discussions; and J. Weston for feedback on early drafts of this paper. **Funding:** All funding was provided by Meta. **Author contributions:** Authors are listed alphabetically in the byline. A.B., N.B., E.D., G.F., C.F., D.F., J.G., H.H., A.P.J., M.Ko., M.Kw., A.L., M.L., A.R., S.R., W.S., A.W., D.W., and H.Z. contributed to the development of Cicero algorithms, code, and experiments. A.G., K.K., and M.Z. provided *Diplomacy* expertise and data annotation. S.M. and D.R. contributed to data collection. A.H.M. and J.S. managed the research team. A.B., N.B., E.D., G.F., C.F., D.F., J.G., A.P.J., A.L., M.L., A.H.M., A.R., W.S., A.W., and D.W. wrote the paper. **Competing interests:** None declared. **Data and materials availability:** The figure and table data are deposited in Zenodo (*33*). The codebase is available at (*34*). Training data was licensed from WebDiplomacy.net by Meta AI.

License information: Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.ade9097
Materials and Methods
Figs. S1 to S10
Tables S1 to S15
References (*35–90*)

Submitted 15 September 2022; accepted 9 November 2022
10.1126/science.ade9097

GEOLOGY

A new biologic paleoaltimetry indicating Late Miocene rapid uplift of northern Tibet Plateau

Yunfa Miao[1,2,3]*, Xiaomin Fang[3]*, Jimin Sun[4], Wenjiao Xiao[5,2], Yongheng Yang[1,2], Xuelian Wang[1,2], Alex Farnsworth[6,3], Kangyou Huang[7]*, Yulong Ren[8], Fuli Wu[3], Qingqing Qiao[5,2], Weilin Zhang[3], Qingquan Meng[9], Xiaoli Yan[9], Zhuo Zheng[7], Chunhui Song[9], Torsten Utescher[10]

The uplift of the Tibet Plateau (TP) during the Miocene is crucial to understanding the evolution of Asian monsoon regimes and alpine biodiversity. However, the northern Tibet Plateau (NTP) remains poorly investigated. We use pollen records of montane conifers (*Tsuga*, *Podocarpus*, *Abies*, and *Picea*) as a new paleoaltimetry to construct two parallel midrange paleoelevation sequences in the NTP at 1332 ± 189 m and 433 ± 189 m, respectively, during the Middle Miocene [~15 million years ago (Ma)]. Both midranges increased rapidly to 3685 ± 87 m in the Late Miocene (~11 Ma) in the east, and to 3589 ± 62 m at ~7 Ma in the west. Our estimated rises in the east and west parts of the NTP during 15 to 7 Ma, together with data from other TP regions, indicate that during the Late Miocene the entire plateau may have reached a high elevation close to that of today, with consequent impacts on atmospheric precipitation and alpine biodiversity.

The Tibet Plateau (TP) is the world's highest inhabited terrain, with an area >2.5×10^6 km^2 and an average elevation >4000 meters above sea level (m asl); it is also commonly referred to as the "Third Pole" or the "Asian Water Tower". Its uplift is a direct expression of the geodynamic processes of Earth's interior caused by the collision of the India-Asia plates (*1–3*) and regarded as the main factor driving the evolution of the Asian monsoon regime (*3–6*) and alpine biodiversity (*7–9*). Despite considerable controversy (*10*, *11*), the Miocene epoch is now deemed critical for understanding the timing of the attainment of the highest paleoelevation of the TP (*12–15*), the development of the modern Asia monsoon regime (*3–6*), and the acceleration of the evolution of alpine biodiversity (*7*, *8*). However, compared with the abundant proxy data and paleoelevation reconstructions for the southern TP (e.g., based on mammalian ecology (*16*), plant fossils (*15*), sedimentary characteristics including isotope geochemistry (*17*), and tectonic collapse) (*12*), the northern part—despite being an integral part of the TP—remains poorly understood.

Tectonic and sedimentary evidence has shown that the NTP has generally been increasing in height since the Middle-Late Miocene (*18–23*),

[1]Key Laboratory of Desert and Desertification, Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou 730000, China. [2]National Key Laboratory of Arid Area Ecological Security and Sustainable Development, Chinese Academy of Sciences, Urumqi 830011, China. [3]State Key Laboratory of Tibetan Plateau Earth System Science, Resources and Environment (TPESRE), Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing 100101, China. [4]Key Laboratory of Cenozoic Geology and Environment, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing 100029, China. [5]Xinjiang Research Center for Mineral Resources, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi 830011, China. [6]School of Geographical Sciences and Cabot Institute, University of Bristol, Bristol BS8 1SS, UK. [7]Guangdong Provincial Key Laboratory of Geodynamics and Geohazards, School of Earth Sciences and Engineering, Sun Yat-sen University, Zhuhai 519082, China. [8]Key Laboratory of Arid Climate Change and Disaster Reduction of CMA and of Gansu Province, Institute of Arid Meteorology, Lanzhou 730000, China. [9]School of Earth Sciences and Key Laboratory of Mineral Resources in Western China (Gansu Province), Lanzhou University, Lanzhou 730000, China. [10]Senckenberg Research Institute, Frankfurt am Main, Steinmann Institute, Bonn University, 53115 Bonn, Germany.
***Corresponding author. Email: miaoyunfa@lzb.ac.cn (Y.M.); fangxm@itpcas.ac.cn (X.F.); hkangy@mail.sysu.edu.cn (K.H.)**

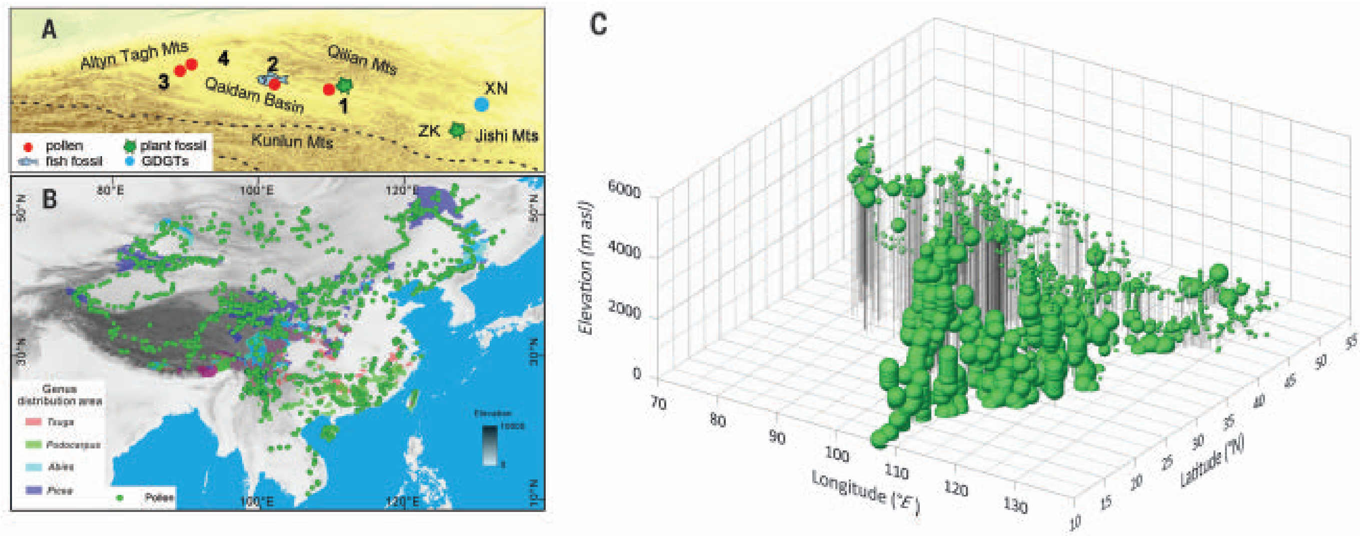


Fig. 1. Locations of sites used for pollen-based paleoelevation reconstruction in the NTP and the distribution of surface pollen samples in Asia. (**A**) 1-HTTL section, 2-YH section, 3-KC-1 core, and 4-SG-1 core used in this study and other sites referenced in the adjunct Qaidam Basin: XN (Xining) and ZK (Zeku); GDGTs (branched glycerol dialkyl glycerol tetraethers); (**B**) Distributions of the four conifer genera and the 3088 surface pollen samples; (**C**) [*Tsuga* (%) + *Podocarpus* (%)] / [*Tsuga* (%) + *Podocarpus* (%) + *Abies* (%) + *Picea* (%)] (TP/TPAP) ratios across the Central East Asia. The base map is a Digital Elevation Model (DEM).

but much older (Oligocene or Eocene) (*24*, *25*) or younger (Late Pliocene) (*26*) uplifts of the region have recently been reported. There is also no consensus on when and how such rapid uplift began to form present range-basin configurations (*18–20*, *23*, *25–27*). Paleoelevation reconstruction can provide key constraints to clarify this question. However, a few current estimations of the elevation history of the NTP from isotope paleoelevation and biology indicate a high discrepancy (*16*, *27–30*). Hydrogen isotope composition and glycerol dialkyl glycerol tetraethers (GDGTs)-based paleotemperature history derived from organic geochemistry demonstrated that the NTP experienced ~2100 to 1000 m of rapid uplift during the Middle to Late Miocene (*27–29*), whereas the oxygen isotope data from paleosol carbonate nodules and Climate-Leaf Analysis Multivariate Program (CLAMP) analysis of fossil leaves in the Qaidam Basin in the central NTP showed that the basin rapidly increased in elevation from ~2000 m at 52 to 44 Ma (*31*) to its present elevation at ~3300 m at ~30.8 Ma (*32*). In addition to the difficulties resulting from large uncertainties rooted in use of isotope-based paleoelevation (*10*, *33*, *34*), biological evidence from fossil mammals, leaves, and pollen in the NTP revealed a very different uplift scenario, all demonstrating much lower elevations and an absence of high Mountains surrounding the Qaidam Basin during the Miocene (*16*, *35*). This inhibits a comprehensive understanding of the paleoelevation history of the entire TP and obscures its role in the evolution of atmospheric circulation (precipitation) or biodiversity.

Montane conifers only grow in higher elevation regions of Earth's surface (*36*, *37*) and consequently their growth and reproduction are sensitive to changes in elevation (*38*, *39*). We focus on the pollen records of four conifer genera (*Tsuga*, *Podocarpus*, *Abies*, and *Picea*) with easily identifiable vesiculate palynomorphology, obtained from sedimentary sequences in the NTP. Our aims were to reconstruct two parallel elevation histories since 16 Ma, produce a geodynamical end-element model of uplift, and link the paleoelevation records from across the TP, to reveal the impacts of uplift on climatic and biological evolution.

The Qaidam Basin covers an area of ~2.5 × 10^5 km^2 with an average elevation of 3000 ± 200 m asl. It is surrounded by three mountain ranges: the Kunlun, Altyn Tagh, and Qilian, which have elevations of ~4000 to 5500 m asl (Fig. 1A). During the Cenozoic the basin accumulated enormous quantities of clastic sediments derived from the adjacent Mountains In the eastern basin, Site 1 [Huaitoutala (HTTL) section] (37°13′ N, 96°43′ E, 2950 m asl) has magnetostratigraphic ages of Middle Miocene to Pliocene (15.6 to 4.0 Ma) (*18*), whereas Site 2 [Yahu (YH) section] (37°46′ N, 93°36′ E, 2838 m asl) is dated to the Pliocene to Early Pleistocene (5.3 to 1.8 Ma) (*40*). We collectively name these two sites East Qaidam. In the western basin, Site 3 (core KC-1) (38°03′ N, 91°45′ E, 2820 m asl), dated to 18 to 5 Ma (*41*), and Site 4 (core SG-1 series) (38°24′ N, 92°30′ E, 2900 m asl), dated to 7 to 0 Ma (*42*), are contemporaneous parallel records, and we collectively name them West Qaidam (Fig. 1A). The maximum distance between these eastern and western sites is >400 km.

At Site 1 we conducted pollen analyses on 34 samples which almost met the statistical requirements of a total >300 palynomorphs [see Sections 1.1 and 1.2 in (*43*)]. Additionally, we used previously published pollen data for 214, 48, and 707 samples from Sites 2 (*40*), 3 (*41*), and 4 (*42*), respectively, which were analyzed by our methodology (Fig. 2, A to D).

Tsuga, *Podocarpus*, *Abies*, and *Picea* are four montane conifer genera represented in the pollen assemblages. Although influenced by factors such as precipitation, geographical location, and soil properties among others, these tree taxa have elevation-specific habitats in Asia (Figs. 1B and 3A): *Tsuga* and *Podocarpus* grow at middle elevations (300 to 3500 and 600 to 2000 m asl, respectively) whereas *Abies* and *Picea* grow at higher elevations (500 to 4700 and 300 to 4700 m asl, respectively) (*36*, *37*). Therefore, the representation of these genera in pollen assemblages can be used to reconstruct paleoelevation. To do this, we first calculated the ratios of [*Tsuga* (%) + *Podocarpus* (%)] to [*Tsuga* (%) + *Podocarpus* (%) + *Abies* (%) + *Picea* (%)] (abbreviated to the TP/TPAP ratio) at 1-Ma intervals to reduce the influence of differences in sample resolution and to evaluate their relationships with climatic fluctuations. The results show that both TP/TPAP records decrease continuously, with the fitting of third-degree polynomial curves resulting in coefficient of determination (R^2) values of 0.94 and 0.90 for East and West Qaidam, respectively (fig. S3C). We also selected steppe taxa (e.g., *Ephedra*, *Nitraria*, and Amaranthaceae) to represent changes in dryness. The results show that in East Qaidam the minimal steppe contents occurred at ~8 Ma, whereas in West Qaidam there was a continuous intensification of aridification (Fig. 2E). Notably, the TP/TPAP ratios are inconsistent with the records

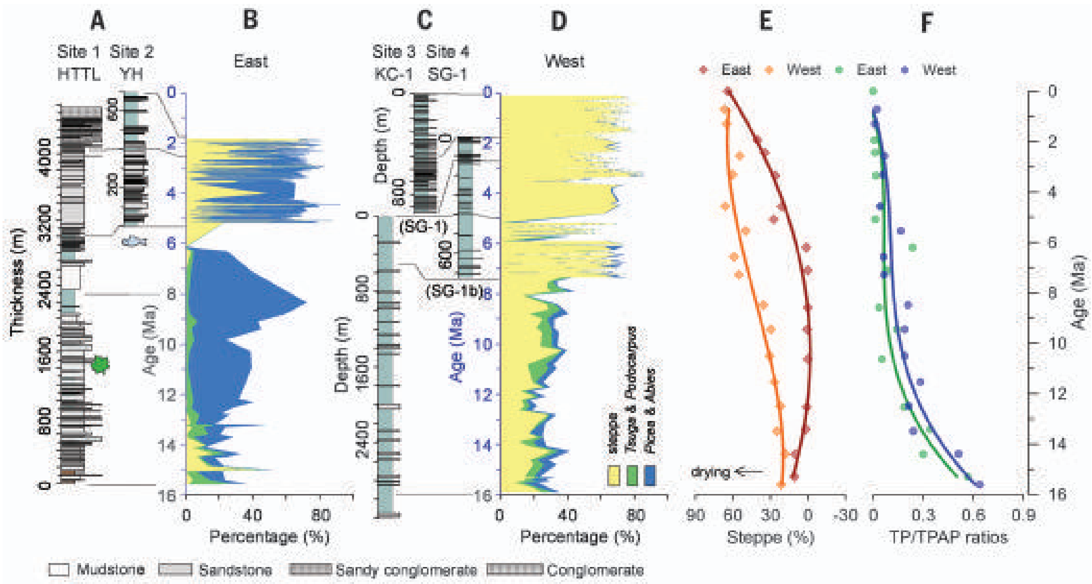


Fig. 2. Lithology, ages, and simplified pollen assemblages. (**A**) and (**B**) Sites 1 and 2 in East Qaidam, respectively. (**C**) and (**D**) Sites 3 and 4 in West Qaidam, respectively. Site 1 is a new pollen study and sites 2 to 4 are redrawn from references (*40*–*42*). (**E**) Steppe taxa percentages and (**F**) TP/TPAP ratios for East and West Qaidam at 1-Ma intervals. Green leaf symbol, plant fossil horizon discovered in this study; fish symbol, horizon of the Cyprinid *Hsianwenia wui* referenced in this study (*46*).

of tetraether temperatures of soil and lacustrine sediments from East Qaidam (*28*), which are based on 9 soil and 15 lacustrine facies samples, with an average temperature deviation between them of 4.3 to 4.6°C, and which reveals a complicated pattern of variation (fig. S3, D to E). Furthermore, the TP/TPAP ratios also differ from that of the global surface temperature (*44*), which shows a lower amplitude of change weaker than that of the TP/TPAP ratios, especially during 16 to 8 Ma (figs. S3, D and E, and S10).

Paleoelevation verification and results

After evaluating the influences of atmospheric and fluvial dynamics, precipitation, temperature, and CO_2 on the TP/TPAP [Sections 1.5 and 1.6 in (*43*)], the modern TP/TPAP–elevation relationship is the foundation of our elevation reconstruction, and our methodology is as follows: First, we determined the modern spatial distribution of the TP/TPAP ratios, based on 3088 surface pollen samples from Asia. The results indicated that most of the samples with high values are from the lower mountain areas in southeastern Asia (Fig. 1C). Second, we determined linear relationships between the TP/TPAP ratio and elevation. The corresponding R^2 value reached 0.90, after correcting the samples to a specific latitude (here, 37.2° N), and then calculating the arithmetic mean for elevation bins at 200-m intervals [Section 1.4 in (*43*), and Fig. 3B]. Third, to test the reliability of the relationships, we calculated the elevations of a series of pollen sites from across the TP (the Himalayas and its deep interior) (five sites of Quaternary age and five sites of Miocene-Pliocene age), together with one site in Japan near sea level [Section 1.9 in (*43*)]. The results for the younger interval are consistent with the middle altitudinal range (midrange) of the drainage (R^2 = 0.80) (table S7). Additionally, the results for the older interval, after correction for the temperature difference (T_D, paleo and modern at sea level, table S5) and the terrestrial temperature lapse rates (TLR) (table S6) using more reliable wet-bulb derived terrestrial lapse rates (*45*) [Section 1.7 in (*43*)], are well matched with records based on isotopes, fossil species with the nearest living relative, and on the CLAMP (table S11). In the NTP, our reconstructed paleoelevation record shows that the parallel midrange paleoelevation of the area adjacent to East Qaidam rose rapidly, from 1332 ± 189 to 3685 ± 87 m asl during the interval of ~15 to ~11 Ma (an increase of ~2400 m) (Fig. 4A and table S9). The region adjacent to West Qaidam also rose rapidly, from 433 ± 189 to 3589 ± 62 m asl during the interval from ~15 to ~7 Ma (an increase of ~3200 m) (Fig. 4A and table S9). Fourth, to determine the basin and mountain elevations separately, we also estimated the basin paleoelevation of the Qaidam Basin using the pollen-based Coexistence Approach during 16 to 14 Ma and the plant fossil–based CLAMP during ~11 Ma, which indicate basin paleoelevations of 1102 ± 429 and 2355 ± 703 m asl, respectively (Fig. 4, B and C). We also cite an estimate of the basin's paleoelevation of ~2362 ± 62 m asl, based on the occurrence of the Cyprinid *Hsianwenia wui* gen. et sp. nov. from the YH section (Site 2) (Late Miocene to earliest Pliocene) (*46*, *47*) [Sections 1.8 and 1.9 in (*43*)], indicating a ~1300 m increase. Thus, we estimate that the montane paleoelevation of the region adjacent to East Qaidam increased from ~1400 to 4200 m asl, and then to 5023 m asl at ~15, ~11, and ~7 Ma, respectively, corresponding to respective increases of ~2800 and ~800 m. In summary, the NTP may have experienced very strong uplifts during the Middle to Late Miocene and attained an elevation similar to its present one during the Late Miocene (8 to 6 Ma).

Rise of the NTP

Relatively few paleoelevation reconstructions exist for the basin and Mountains in the NTP, and the available reconstructions have large potential uncertainties depending on the different proxies and methods used (*28*, *29*, *48*). Tetraether temperatures revealed a relatively substantial relief (>700 to 1100 m) between the Qaidam Basin and the Qilian Mountains since ~12.4 Ma (*28*). A multidisciplinary approach incorporating regional zircon geochronology indicated that the Jishi Mountains, to the southeast of the Qilian Mountains, were at a low elevation during 14 to 9 Ma and then rose abruptly at ~8 Ma to reach an elevation sufficient to block the passage of water vapor (*49*). Tectonically, there was the synchronous emergence and rapid erosion of the Qilian Mountains at ~12 Ma (*28*). Thermodynamics-based evidence also exists for the onset of the rapid exhumation of the Qilian Mountains at ~10 Ma (*50*). Our results have quantified the paleoelevation of most of the crustal shortening across the Qilian Mountains that accumulated during the Late Middle to Late Miocene.

For the basin elevation, leaf wax hydrogen isotopes (*28*) indicate that the elevation of the Qaidam Basin increased by ~2100 m during 15 to 10 Ma—almost twice that indicated by our results (a rise of ~1100 m). At the same time, the elevation of the Xining Basin to the east increased rapidly by ~1000 m at ~10 to 8 Ma, according to a record of GDGTs (*29*). Since the present-day elevation of the Xining Basin is ~2200 m asl, the paleoelevation of the basin should have been <1200 m asl at 10 to 8 Ma.

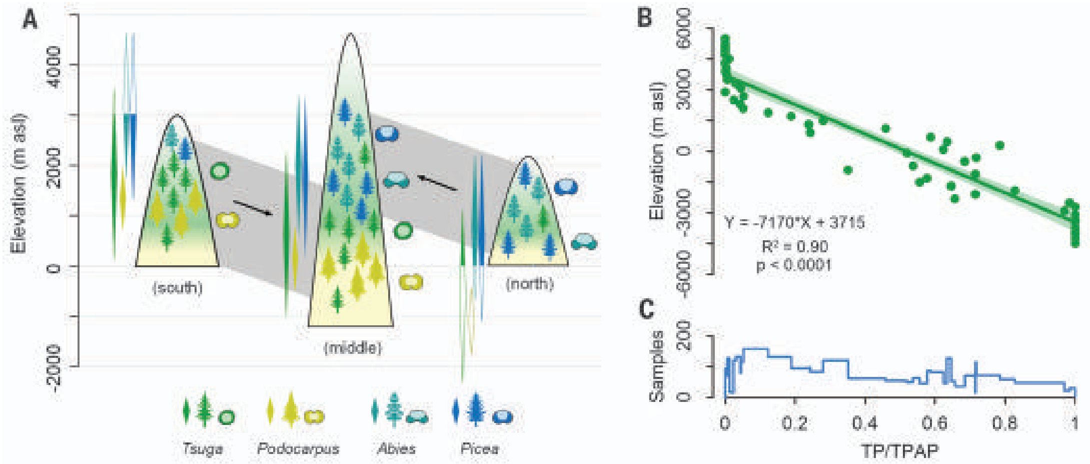
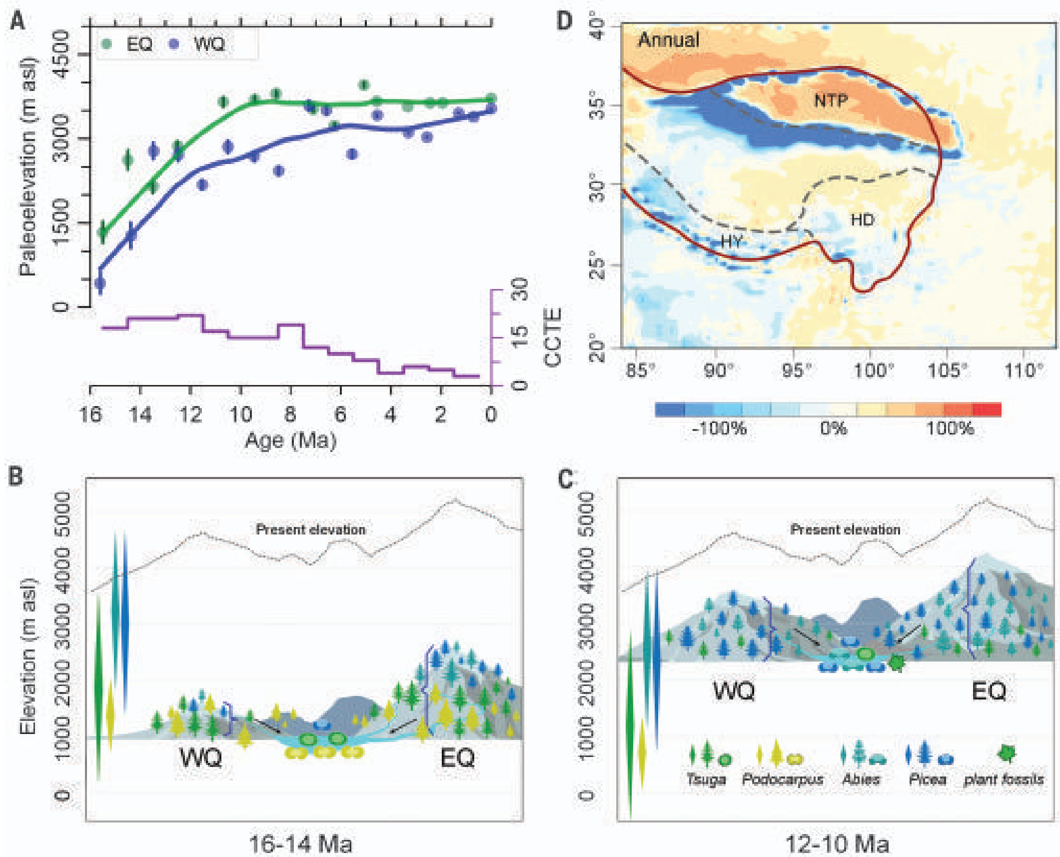


Fig. 3. Elevation of the studied conifer genera corrected for latitude and pollen results versus corrected elevation. (**A**) Sketch map of genera elevations at a specific latitude and results corrected for latitude (arrows show the directions of latitudinal correction). (**B**) Correlation between the TP/TPAP ratios and the elevation at 200 m intervals, after correcting the surface pollen samples to the latitude of 37.2° N (light green shaded area is the 95% confidence interval). (**C**) Sample numbers of surface pollen assemblages for each 200-m height interval.

Fig. 4. NTP uplift and impact on precipitation. (**A**) Pollen-based paleoelevation curves at 1-Ma intervals with standard error of the mean error bars (table S9), with terrestrial temperature lapse rates (Γt_{wb}) of 3.1 to 4.0 °C 1000 m^{-1} (table S6), and green and blue lines show the LOESS Fit (Span = 0.5). Purple line shows a cumulative curve of tectonic event numbers (CCTE) in the NTP (north of the Kunlun Suture). (**B**) and (**C**) Illustrations of the distribution of conifers versus elevation for two time slices, after temperature correction. The black brace and arrow show the midrange of the drainage and the pollen transport direction, respectively. The cyan shading indicates rivers and lakes within the lowest part of the basin. (**D**) RegCM4.6 model results showing the percentage precipitation anomaly after reducing the elevation of the NTP by 67% [Section 1.11 in (*43*)]. The red line shows an outline of the entire TP, and the dashed lines indicate the boundaries of the NTP, Himalayas (HY), and Hengduan Mountains (HD).



Additionally, the elevation of the Zeku Basin (see Fig. 1 for location, present day elevation 3440 to 4300 m asl) during the Middle Miocene is speculated to have been 1200 to 1400 m asl (*51*) or <2400 m asl (*52*). Combining this evidence, we conclude that the basins in the NTP were at lower elevations before the Late Miocene and were then intensely uplifted until 8 to 6 Ma.

Maximum elevation of the entire TP

In the southern TP, there are several north-south rifts and normal faults (fig. S17), including Leo Pargil (15.5 to 14.5 Ma) (*53*), Thakkhola Hydrothermal Muscovite (~14 Ma) (*12*), Shuanghu Graben (~13.5 Ma) (*13*), and Yangbajing Graben (8 Ma) (*3*). Additionally, stable isotope evidence indicates the following elevations and ages: Namling-Oiyug Basin (~15 Ma, 5200 + 1400 to 600 m) (*54*) or (~5400 m) (*55*), Mount Everest (15 Ma, 5100 to 5400 m) (*56*), Zhada Basin

(9.2 Ma, ~5500 to 6500 m) (*14*), Thakkhola (11 to 7 Ma, ~5500 m) (*57*), and the Gyirong Basin (10 Ma, 5800 m) (*58*). These estimates are supported by our pollen-based results, which together indicate that the elevation of the entire TP during the late Middle to early Late Miocene may have been similar to that of today.

These records are consistent with the conclusion that east-west extension caused the elevation of the Kailas and Qiabulin areas in the southern TP to decrease by ~1000 m in the Late Miocene (*59*) whereas the elevation of the Songpan Basin in the eastern TP increased to a height similar to its present day elevation by 11 Ma (*60*). Molnar *et al.* (*61*) proposed a "convective stripping model" to explain the uplift and east-west extension of the plateau, based on the development of north-south normal faults, which was subsequently supported by seismic results (*62*). Overall, the rapid uplift of the Qaidam Basin and Qilian Mountains in the Middle to early Late Miocene was almost simultaneous with the east-west extension of the rifting in the central-southern TP (*61*) and other coeval tectonic events within and around the TP [a summary is given in (*25*)]. These findings support the end-member model of the removal of mantle lithosphere beneath different portions of the TP (*3*, *25*, *61*).

Impact of uplift on climate and biodiversity changes

We propose that the entire TP may have reached its maximum elevation and spatial extent during the Late Miocene, which would have affected regional climatic and biological evolution through physical blocking and thermodynamic effects on the transport of atmospheric water vapor and the resulting precipitation (*3*, *6*, *63*). To quantitatively evaluate the precipitation variability in the TP, we ran the Regional Climate Model (RegCM4.6) with a 30-km resolution, reducing the elevation of the NTP to 33% in a sensitivity test (similar to the scenario of a low TP during the Middle Miocene, called exp. LM). By contrast, a control test was run which maintained the present-day topography of the TP, called exp. HP [Section 1.11 in (*43*)]. The results show that, in contrast to exp. HP, the annual precipitation in the NTP was reduced by >50% in exp. LM (Fig. 4D) and was slightly higher in the spring and summer than in the autumn and winter (fig. S21). However, in the Himalayas and Hengduan Mountains to the south and southeast, the precipitation increased by 50% and 150%, respectively (Fig. 4D). These precipitation changes are supported by the results of the RegCM4.1 (*64*), the Community Atmosphere Model (CAM4) (*65*), and the Weather Research and Forecasting model, nested within CAM4 (*66*), which was used to model the removal of the NTP as well as other marginal regions surrounding the TP. The principal reasons for the precipitation changes are thought to be the strengthening of the East Asian summer monsoon circulation (*65*) and a decrease in local precipitation in the Asian interior, driven by the enhanced rain shadow effect of the Mountains and changes in regional circulation (*64*). Progressive aridification in the west is evidenced by the increase in steppe taxa in pollen assemblages (Fig. 2E) and occurrence of the unusual thick-boned Cyprinid, *Hsianwenia wui* (*46*) during the Late Miocene to Early Pliocene, whereas in the east a wetting-drying transition occurred at ~8 Ma (Fig. 2E). Despite this spatial inconsistency, both *Podocarpus* and *Tsuga* flourished in areas of low topography and then decreased slightly until 8 to 6 Ma, although *Picea* and *Abies* also flourished in the east (Fig. 2B) and decreased in the west (Fig. 2D), and the trends of the TP/TPAP ratios are synchronous (Fig. 2F).

The pronounced precipitation increase in the Himalayas and Hengduan Mountains is also correlative with increases in the diversity of vegetation (*8*) and reptiles (*7*) in these regions during the Middle to Late Miocene, against the background of a high TP. Thus, we suggest that these increases in biodiversity were driven jointly by mountain building and intensification of the Asian monsoon (*8*, *67*) and that overall, the uplift of the NTP can be seen to have contributed substantially to climate change and biological evolution in Asia.

We used two parallel montane conifer pollen records from the NTP to reconstruct paleoelevation records for the interval since the Middle Miocene. Our results clearly show that the high elevation in the NTP was attained at ~10 to 7 Ma, supporting the geodynamics model of the removal of mantle lithosphere beneath different portions of the TP. The rise of the NTP during the Miocene may also have enhanced the role of topography in accelerating the rate of change of regional biodiversity through the influence of topography on precipitation.

REFERENCES AND NOTES

1. P. Tapponnier *et al.*, *Science* **294**, 1671–1677 (2001).
2. C. Wang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 4987–4992 (2008).
3. T. M. Harrison, P. Copeland, W. S. Kidd, A. Yin, *Science* **255**, 1663–1670 (1992).
4. P. Molnar, W. R. Boos, D. S. Battisti, *Annu. Rev. Earth Planet. Sci.* **38**, 77–102 (2010).
5. Z. An, J. E. Kutzbach, W. L. Prell, S. C. Porter, *Nature* **411**, 62–66 (2001).
6. F. Wu *et al.*, *Nat. Rev. Earth Env.* **3**, 684–700 (2022).
7. W. Xu *et al.*, *Natl. Sci. Rev.* **8**, nwaa263 (2020).
8. W.-N. Ding, R. H. Ree, R. A. Spicer, Y.-W. Xing, *Science* **369**, 578–581 (2020).
9. S. F. Li *et al.*, *Sci. Adv.* **7**, eabc7741 (2021).
10. S. Botsyun *et al.*, *Science* **363**, eaaq1436 (2019).
11. A. Licht *et al.*, *Nature* **513**, 501–506 (2014).
12. M. Coleman, K. Hodges, *Nature* **374**, 49–52 (1995).
13. P. M. Blisniuk *et al.*, *Nature* **412**, 628–632 (2001).
14. J. E. Saylor *et al.*, *Am. J. Sci.* **309**, 1–42 (2009).
15. R. A. Spicer *et al.*, *Nature* **421**, 622–624 (2003).
16. T. Deng, L. Ding, *Natl. Sci. Rev.* **2**, 417–437 (2015).
17. D. B. Rowley, B. S. Currie, *Nature* **439**, 677–681 (2006).
18. X. Fang *et al.*, *Earth Planet. Sci. Lett.* **258**, 293–306 (2007).
19. J. Li *et al.*, *Quat. Res.* **81**, 400–423 (2017).
20. R. O. Lease, D. W. Burbank, B. Hough, Z. C. Wang, D. Y. Yuan, *Geol. Soc. Am. Bull.* **124**, 657–677 (2012).
21. B. Li *et al.*, *J. Geophys. Res. Solid Earth* **122**, 4106–4122 (2017).
22. D. Zheng *et al.*, *Lithosphere* **9**, 408–416 (2017).
23. X. Fang *et al.*, *Sci. China. Ser. D Earth Sci.* **45**, 289–299 (2002).
24. H. Lu *et al.*, *Geol. Soc. Am. Bull.* **134**, 2531–2547 (2022).
25. W. Wang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2120364119 (2022).
26. F. Richter *et al.*, *Geol. Soc. Am. Bull.*B36302.1 (2022).
27. G. Zhuang, M. T. Brandon, M. Pagani, S. Krishnan, *Earth Planet. Sci. Lett.* **390**, 186–198 (2014).
28. G. Zhuang *et al.*, *Geophys. Res. Lett.* **46**, 1312–1319 (2019).
29. C. Chen *et al.*, *Geophys. Res. Lett.* **46**, 8375–8386 (2019).
30. X. Wang *et al.*, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **254**, 363–385 (2007).
31. B. Song *et al.*, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **512**, 6–22 (2018).
32. B. Song *et al.*, *Earth Planet. Sci. Lett.* **537**, 116175 (2020).
33. J. Quade *et al.*, *Global Planet. Change* **191**, 103194 (2020).
34. Z. Jing *et al.*, *Nat. Commun.* **13**, 4371 (2022).
35. X. Wang *et al.*, *Earth Planet. Sci. Lett.* **208**, 253–269 (2003).
36. A. Farjon, D. Filer, *An Atlas of the World's Conifers: An Analysis of Their Distribution, Biogeography, Diversity and Conservation Status*. (Brill, 2013), pp. 171–292.
37. L. Fu, T. Chen, K. Lang, T. Hong, Q. Lin, *Higher Plants of China, Vol. 3*. (Qingdao Publishing House, 2000).
38. H. Lu *et al.*, *Boreas* **37**, 254–262 (2008).
39. G. Dupont-Nivet, C. Hoorn, M. Konert, *Geology* **36**, 987–990 (2008).
40. F. Wu, X. Fang, M. Herrmann, V. Mosbrugger, Y. Miao, *Global Planet. Change* **76**, 16–21 (2011).
41. Y. Miao *et al.*, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **299**, 30–38 (2011).
42. A. Koutsodendris *et al.*, *Global Planet. Change* **180**, 117–125 (2019).
43. Materials and methods are available as supplementary materials.
44. T. Westerhold *et al.*, *Science* **369**, 1383–1387 (2020).
45. A. Farnsworth *et al.*, *Earth Planet. Sci. Lett.* **564**, 116903 (2021).
46. M. Chang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 13246–13251 (2008).
47. M. Chang, D. Miao, *Kexue Tongbao* **61**, 981–995 (2016).
48. M. L. Kent-Corson *et al.*, *Earth Planet. Sci. Lett.* **282**, 158–166 (2009).
49. J. E. Saylor *et al.*, *Basin Res.* **30**, 544–563 (2018).
50. D. Zheng, M. K. Clark, P. Zhang, W. Zheng, K. A. Farley, *Geosphere* **6**, 937–941 (2010).
51. Z. Hui *et al.*, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **511**, 419–432 (2018).
52. X. Li *et al.*, *J. Syst. Evol.* **59**, 1321–1338 (2021).
53. R. C. Thiede *et al.*, *Geol. Soc. Am. Bull.* **118**, 635–650 (2006).
54. B. S. Currie, D. B. Rowley, N. J. Tabor, *Geology* **33**, 181–184 (2005).
55. M. A. Khan *et al.*, *Global Planet. Change* **113**, 1–10 (2014).
56. A. Gébelin *et al.*, *Geology* **41**, 799–802 (2013).
57. C. N. Garzione, D. L. Dettman, J. Quade, P. G. DeCelles, R. F. Butler, *Geology* **28**, 339–342 (2000).
58. D. B. Rowley, R. T. Pierrehumbert, B. S. Currie, *Earth Planet. Sci. Lett.* **188**, 253–268 (2001).
59. Q. Xu *et al.*, *Gondwana Res.* **54**, 50–61 (2018).
60. Q. Xu, X. Liu, L. Ding, *Geochem. Geophys. Geosyst.* **17**, 4254–4267 (2016).
61. P. Molnar, P. England, J. Martinod, *Rev. Geophys.* **31**, 357–396 (1993).
62. L. Chen, F. A. Capitanio, L. Liu, T. V. Gerya, *Nat. Commun.* **8**, 15992 (2017).
63. G. Wu *et al.*, *Sci. Rep.* **2**, 404 (2012).
64. X. Liu, H. Sun, Y. Miao, B. Dong, Z.-Y. Yin, *Quat. Sci. Rev.* **116**, 1–14 (2015).
65. R. Zhang, D. Jiang, X. Liu, Z. Tian, *Chin. Sci. Bull.* **57**, 4617–4626 (2012).
66. E. Yu *et al.*, *Global Planet. Change* **169**, 34–47 (2018).
67. A. Antonelli *et al.*, *Nat. Geosci.* **11**, 718–725 (2018).

ACKNOWLEDGMENTS

We thank H. Yan, Y. Duan, and F. Zhang for field and laboratory assistance, Y. Zhao and X. Xiao for assistance with compiling the data, and all contributors of the present database and all members of the

China Quaternary Pollen Database (CPD) organized in 1998. We also thank R. Spicer and G. Shi for help with fossil identification and CLAMP analysis, F. Chen, X. Zhang, X. Cao, and J. Yang for their valuable discussions and comments to improve the revised manuscript, S. Botsyun and other three referees for constructive suggestions. **Funding:** This research was provided by Second Tibetan Plateau Scientific Expedition program (2019QZKK0707), NSFC (41888101, 41988101, 41772181, 42122028, 42202029), Strategic Priority Research Program of CAS (XDA20070202), National Key R&D Program of China (2022YFF0801501), Chinese Academy of Sciences Visiting Professorship for Senior International Scientists (2021FSE0001), Leverhulme grants (RPG-2019-365; RPG-2021-202), and NERC (NE/X015505/1). **Author contributions:** Y.F.M., X.M.F., Y.H.Y., and K.Y.H. designed the project and developed the conceptual framework. Y.F.M. identified the pollen. Y.F.M., K.Y.H., Y.H.Y., X.L.W., and F.L.W. analyzed the data. Y.F.M., Y.H.Y., Q.Q.Q., K.Y.H., and X.M.F. wrote and revised the manuscript. A.F. carried out the HadCM3 simulations. All authors provided comments and revisions. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the Supplementary Materials. **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.sciencemag.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS
science.org/doi/10.1126/science.abo2475
Materials and Methods
Figs. S1 to S23
Tables S1 to S14
References (*68–199*)

Submitted 22 January 2022; accepted 8 November 2022
10.1126/science.abo2475

EVOLUTIONARY ECOLOGY

Rapid weed adaptation and range expansion in response to agriculture over the past two centuries

Julia M. Kreiner[1,2]*, Sergio M. Latorre[3,4], Hernán A. Burbano[3,4], John R. Stinchcombe[5], Sarah P. Otto[2,6], Detlef Weigel[4], Stephen I. Wright[5]

North America has experienced a massive increase in cropland use since 1800, accompanied more recently by the intensification of agricultural practices. Through genome analysis of present-day and historical samples spanning environments over the past two centuries, we studied the effect of these changes in farming on the extent and tempo of evolution across the native range of the common waterhemp (*Amaranthus tuberculatus*), a now pervasive agricultural weed. Modern agriculture has imposed strengths of selection rarely observed in the wild, with notable shifts in allele frequency trajectories since agricultural intensification in the 1960s. An evolutionary response to this extreme selection was facilitated by a concurrent human-mediated range shift. By reshaping genome-wide diversity across the landscape, agriculture has driven the success of this weed in the 21st century.

Agricultural practices across North America have rapidly intensified over the past two centuries through cropland expansion (*1*), monoculture plantings (*2*, *3*), and increased chemical inputs (*4*, *5*). Since the beginning of the 1800s, cropland usage has expanded from 8 million to 200 million hectares in Canada and the United States alone (*1*). Since the mid-1900s, development of new crop varieties—including high-yield and herbicide-resistant wheat, corn, and soy (*6*, *7*)—has greatly improved the efficiency of food production in all farming sectors. Combined with increased reliance on pesticides, fertilizers, irrigation, and large-scale mechanization, this global transformation is often referred to as the agricultural Green Revolution (*8*–*10*). Pesticide effectiveness, however, has been limited by the evolution of resistance across numerous pest species (*11*–*14*). Although technological innovation for efficient food production has risen with increasing global food demands, the concomitant landscape conversion has become one of the foremost drivers of global biodiversity loss (*15*).

[1]Department of Botany, University of British Columbia, Vancouver, BC, Canada. [2]Biodiversity Research Centre, University of British Columbia, Vancouver, BC, Canada. [3]Centre for Life's Origins and Evolution, Department of Genetics, Evolution and Environment, University College London, London, UK. [4]Department of Molecular Biology, Max Planck Institute for Biology Tübingen, Tübingen, Germany. [5]Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, ON, Canada. [6]Department of Zoology, University of British Columbia, Vancouver, BC, Canada.
***Corresponding author. Email: julia.kreiner@ubc.ca**

Species that have managed to survive, and even thrive, in the face of such extreme environmental change provide notable examples of rapid adaptation on contemporary time scales and illustrate the evolutionary consequences of anthropogenic change. One such species is the common waterhemp (*Amaranthus tuberculatus*), an agricultural weed that is native to North America and persists in large part in natural, riparian habitats (*16*, *17*), which provides an opportunity to investigate the time scale and extent of contemporary agricultural adaptation. The genetic changes underlying weediness are particularly important to understand in *A. tuberculatus* because it has recently become one of the most problematic agricultural weeds in North America as a result of its widespread adaptation to herbicides, its persistence in fields across seasons, and its strong ability to compete with both soy and corn (*18*, *19*). Determining the roles of newly arisen mutations, genetic variants predating the onset of environmental change (*20*, *21*), migration across the range (*22*), and their interactions (*23*, *24*) will inform the temporal and spatial scales at which contemporary adaptation occurs and management strategies should be used.

To understand how changing agricultural practices have shaped the success of a ubiquitous weed, we analyze genomic data from contemporary paired natural and agricultural populations alongside historical herbarium samples collected from 1828 until 2011 (Fig. 1). With this design, we identify candidate agriculturally adaptive alleles (i.e., those that occur at consistently higher frequencies in agricultural compared with nearby natural sites), track their frequencies across nearly two centuries, and link the tempo of weed adaptation to demographic changes and key cultural shifts in modern agriculture.

The genome-wide signatures of agricultural adaptation

To find alleles favored under current farming practices, we looked for those that were consistently overrepresented in extant populations collected in agricultural habitats compared with neighboring riparian (natural) habitats (*25*) using Cochran–Mantel–Haenszel (CMH) tests (Fig. 2A). Alleles associated with agricultural environments [the 0.1% of single-nucleotide polymorphisms (SNPs) with the lowest CMH P values; $n = 7264$] are significantly enriched for 29 gene ontology (GO) biological process terms related to growth and development; reproduction; cellular metabolic processes; and responses to abiotic, endogenous, and external stimuli, including responses to chemicals (table S1). The importance of chemical inputs in shaping weed agricultural adaptation is clear in that the most significant agriculturally associated SNP {raw P value = 8.6×10^{-11}; [false discovery rate (FDR)–corrected] Q value = 0.00062} falls just 80 kb outside the gene protoporphyrinogen oxidase (*PPO*)—the target of PPO-inhibiting herbicides (Fig. 2B). PPO-inhibiting herbicides were widely used in the 1990s and have had a recent resurgence to control and slow the spread of glyphosate-resistant weeds (*26*, *27*). Other genes with the strongest agricultural associations include *ACO1*, which has been shown to confer oxidative stress tolerance (*28*); *HB13*, involved in pollen viability (*29*) as well as drought and salt tolerance (*30*); *PME3*, involved in growth through germination timing (*31*); *CAM1*, a regulator of senescence in response to stress (*32*, *33*); and both *CRY2* and *CPD*, two key regulators of photomorphogenesis and flowering through brassinosteroid signaling (*34*–*37*) (table S2). These changes are consistent with agricultural

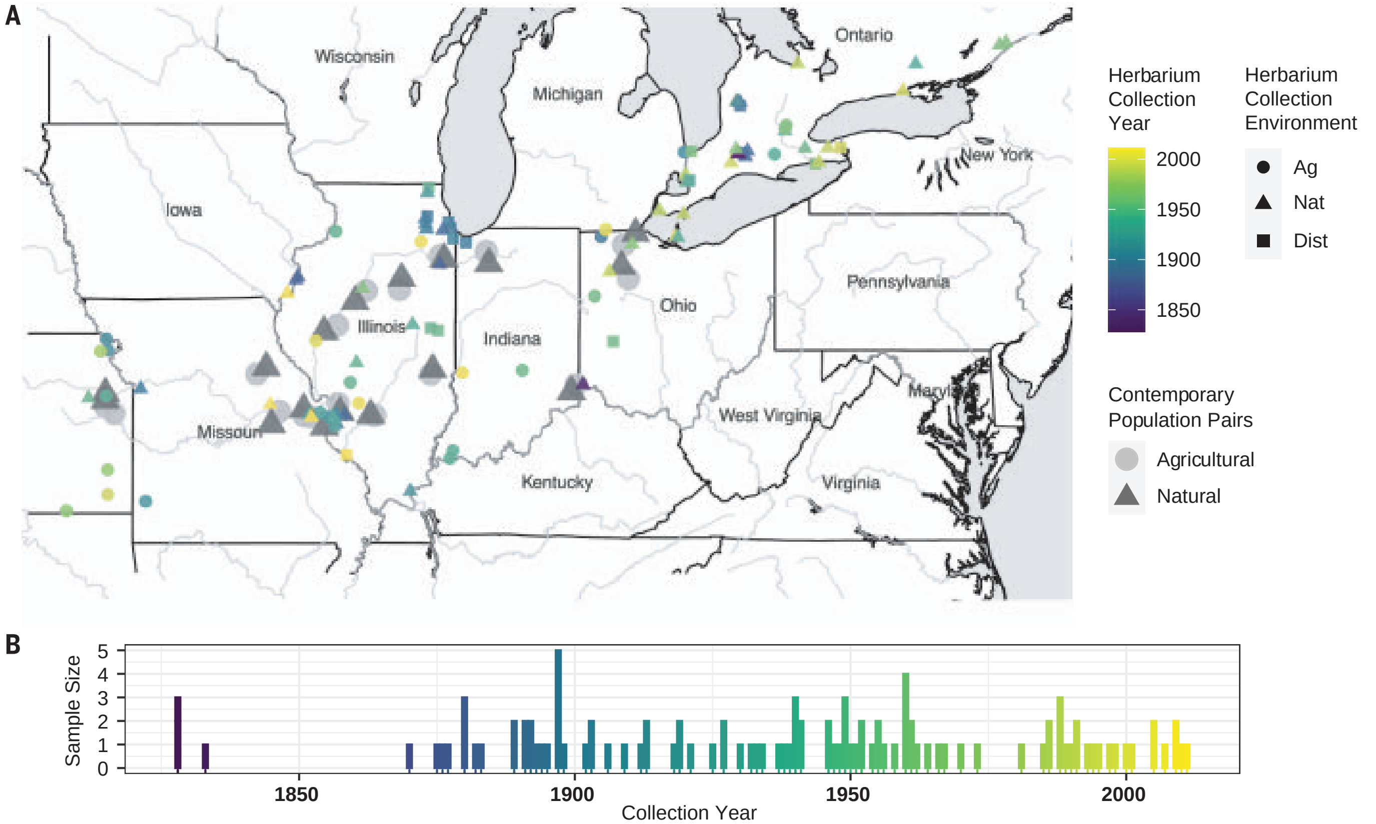


Fig. 1. Sequenced waterhemp collections through space and time. (**A**) Map of 17 contemporary paired natural-agricultural populations [n = 187, collected and sequenced by Kreiner *et al.* (*25*)] along with 108 newly sequenced herbarium specimens dating back to 1828 collected across three environment types: agricultural (Ag), natural (Nat), and disturbed (Dist) (metadata provided in data S2). (**B**) Distribution of sequenced herbarium samples through time.

selection to grow in high-stress and high-disturbance environments among fast-growing crops. Natural-versus-agricultural allele frequency differentiation (F_{ST}) is highly correlated with the CMH test statistic (Pearson's correlation coefficient r = 0.987), with 78% (98%) of CMH focal SNPs overlapping with the top 0.01% (0.1%) of F_{ST} hits (fig. S1). Despite negligible genome-wide differentiation among environments suggesting widespread gene flow (F_{ST} = 0.0008; with even lower mean F_{ST} between paired sites = –0.0029; Fig. 2C), our results suggest that strong antagonistic selection acts to maintain spatial differentiation for particular alleles, with 403 SNPs showing a CMH Q < 0.10 (251 after aggregating linked SNPs).

To further investigate the extent to which herbicides shape adaptation to agriculture, we assayed patterns of environmental differentiation by known herbicide-resistance variants. Eight such alleles were present in contemporary samples, only six of which were common (table S3): a deletion of codon 210 within *PPO* (*38*), a copy number amplification and a nonsynonymous mutation within 5-enolpyruvylshikimate-3-phosphate synthase (*EPSPS*) conferring resistance to glyphosate herbicides (*39*), and 3 separate nonsynonymous mutations within acetolactate synthase (*ALS*) conferring resistance to ALS-inhibiting herbicides (*19*). Although these resistance alleles were at intermediate frequencies in agricultural populations, ranging from 0.08 to 0.35, they tended to be rarer but still frequent in natural populations, ranging from 0.04 to 0.22 (Fig. 2C). Four of six common resistance alleles show significant allele frequency differences across environments (EPSPSamp: F = 9.02, P = 0.008; PPO210: F = 53.39, $P = 1.04 \times 10^{-11}$; ALS574: F = 4.95, P = 0.028; ALS376: F = 4.37, P = 0.038), two of which are among the strongest signals of differentiation genome-wide. Natural-versus-agricultural F_{ST} at the PPO210 deletion, 0.21, is higher than anywhere else in the genome and is even stronger when calculated within population pairs (F_{ST} = 0.27) (Fig. 2C). Similarly, the EPSPS amplification is ranked 20th among genome-wide biallelic F_{ST} values at 0.14 (within-pair F_{ST} = 0.22), in support of herbicides as a foremost driver of agricultural adaptation (Fig. 2D).

To infer the importance of selective trade-offs in adaptation across natural and agricultural environments, we implemented a Wright-Fisher allele frequency–based migration-selection balance model for these four differentiated resistance alleles as well as the top 30 independent CMH outliers. Assuming that these alleles are at a steady state between migration and selection, we inferred that the costs of resistance per migrant that has arrived into natural environments are consistently higher than the benefits of resistance per migrant that has arrived into agricultural environments (per-migrant cost-to-benefit ratio ranges from 1.39 for EPSPSamp to 5.03 for the PPO210 deletion; Fig. 2D and table S3). Thus, the spread of these four common herbicide-resistance alleles appears to be constrained either by more consistent selection against resistance in herbicide-free, natural environments or by particularly high rates of migration of susceptible alleles from natural into agricultural environments. In comparison, for the top 30 independent CMH outliers, the costs per migrant that has arrived in natural environments were about equally likely to be stronger or weaker (12/28, 42%) than the benefits per migrant in agricultural environments (fig. S2). This population genetic inference provides a previously unused and sensitive alternative to experimental studies of fitness costs that vary greatly depending on context (*40*), highlighting the potentially important role of resistance costs across a diverse

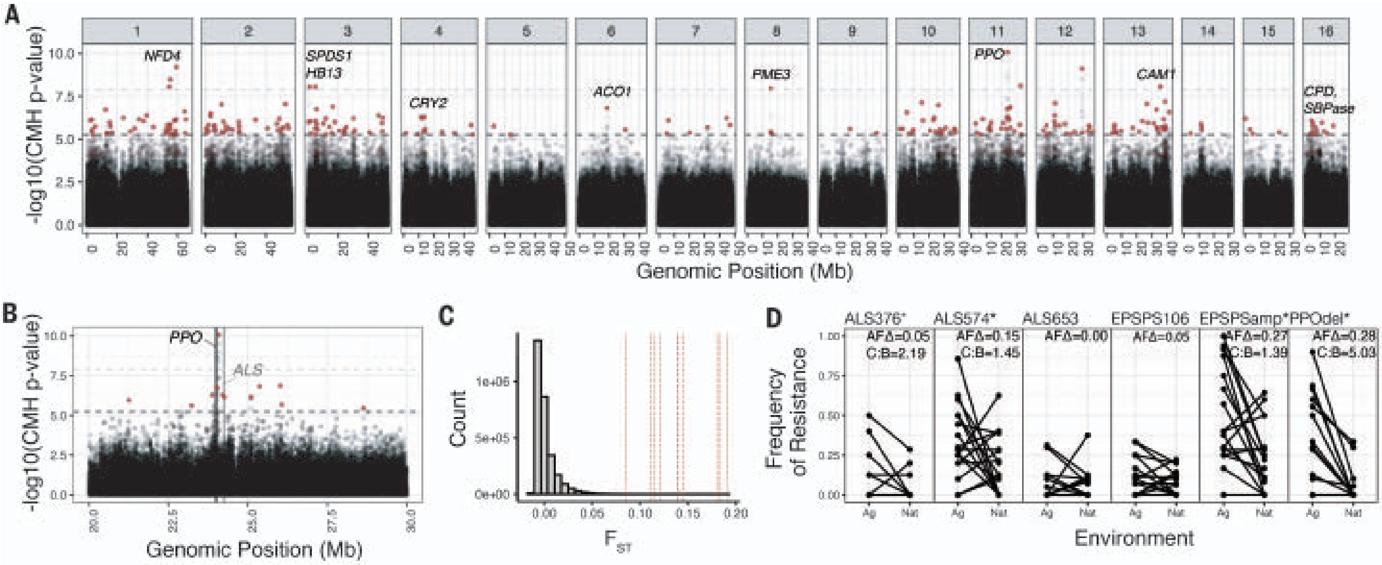


Fig. 2. Signals of contemporary agricultural adaptation, gene flow, and antagonistic selection across the genome in *A. tuberculatus*. (**A**) Results from CMH tests for SNPs with consistent differentiation among environments across contemporary natural-agricultural population pairs. A 10% FDR threshold is indicated by the lower dashed horizontal black line, whereas the Bonferroni-corrected $P < 0.1$ cut-off is shown by the upper dashed horizontal gray line. Red points indicate focal agricultural-associated SNPs after aggregating linked variation ($r^2 > 0.25$ within 1 Mb). Candidate agriculturally adaptive genes for peaks that are significant at a 10% FDR threshold are named. (**B**) CMH results from the scaffold containing the most significant CMH P value, corresponding to variants linked to the PPO210 deletion conferring herbicide resistance and to the nearby herbicide-targeted gene *ALS*. (**C**) Distribution of F_{ST} values between all agricultural and natural samples for ~3 million genome-wide SNPs (minor allele frequency > 0.05). Vertical lines indicate F_{ST} values for the 10 candidate genes named in (A). (**D**) Population-level frequencies of six common herbicide-resistance alleles across geographically paired agricultural and natural habitats sampled in 2018 (pairs connected by horizontal lines). The first four columns are nonsynonymous variants in *ALS* and *EPSPS* followed by EPSPSamp (a 10-Mb-scale amplification that includes *EPSPS*) and an in-frame single-codon deletion in PPO. Estimates of per-migrant natural cost–to–agricultural benefit ratio (C:B) are shown in the top right corner for the four resistance alleles with significant (indicated by asterisks) allele frequency differences (AFΔ) across environment types in a multiple linear regression.

set of individuals within complex agricultural and natural environments. In these field settings, further work is necessary to understand the contributions of temporal and spatial heterogeneity in both migration and selection for and against resistance across the landscape.

Agriculturally adaptive alleles change rapidly

With the genome-wide set of 251 modern agriculture-associated alleles, we searched for signatures of temporal evolution using newly collected whole-genome sequence data from a set of historical herbarium samples ($n = 108$) dating back to 1828. These samples provide snapshots of the genetic changes that have occurred over this time period and across environment types, with collections from natural and weedy (agricultural and disturbed) habitats (Fig. 1). Of the 165 loci for which we had sufficient information in the historical SNP set (sequenced to 10× coverage, on average), 151 were segregating with the same reference-alternate allele combination (i.e., 11 were dropped because of multiallelism), and only three were invariant. To model allele frequency change through time at these alleles, we implemented logistic regressions of genotypes (within-individual allele frequencies) at each locus by collection year, where twice the slope of the logit-transform is equivalent to the strength of selection (s) in a diploid model of selection [where s is the fitness difference between homozygotes, assuming additivity; see materials and methods for model and simulations (*41*)].

Consistent with the rapid change in land use and farming practices in the recent past, the frequency of these 154 contemporary agricultural alleles has increased substantially over the past two centuries. Whereas in natural environments agriculturally associated alleles have increased by 6% on average since 1870—the earliest time point at which we have collections across environment types—these same alleles have increased by 22% in disturbed and agricultural environments (Fig. 3A). This observed change greatly exceeds the expected change over this time period, based on genome-wide patterns that reflect drift, migration, selection, and demographic change [null 95% interquantile range for allele frequency change in natural sites = (–2.7, 2.0.%); for change in agricultural and disturbed sites = (3.3, 7.9%)]. We generated these null expectations by randomly sampling a set of 154 loci with the same distribution of contemporary allele frequencies (fig. S4) and calculating their frequency change through time across herbarium samples, separately in each environment, 1000 times [see materials and methods (*41*)]. That the observed change in natural environments is also more extreme than what is expected is consistent with ongoing migration of agriculturally selected alleles and subsequent costs in natural environments.

The considerable increase in frequency of these alleles across environments corresponds to notably strong selection, even when estimated over century-long time periods. The 154 agriculture-associated alleles collectively exhibit a selective strength of $\tilde{s} = 0.022$ since the 1870s in agricultural and disturbed habitats. However, these alleles exhibit much weaker selection, $\tilde{s} = 0.0056$, in natural habitats [agricultural and disturbed null interquantile range = (0.0026, 0.0068); natural null interquantile range = (–0.0018, 0.0018)]. An open question in evolutionary biology is what distribution of selection coefficients underlie adaptation (*42*). We estimate that selection on agricultural-associated loci varies between –0.196 and 0.150 in natural habitats and between –0.090 and 0.372 in agricultural and disturbed habitats, reflective of left- and right-skewed distributions, respectively (Fig. 3B and fig. S5). The top 15 agriculture-associated alleles that we infer have experienced the strongest selection

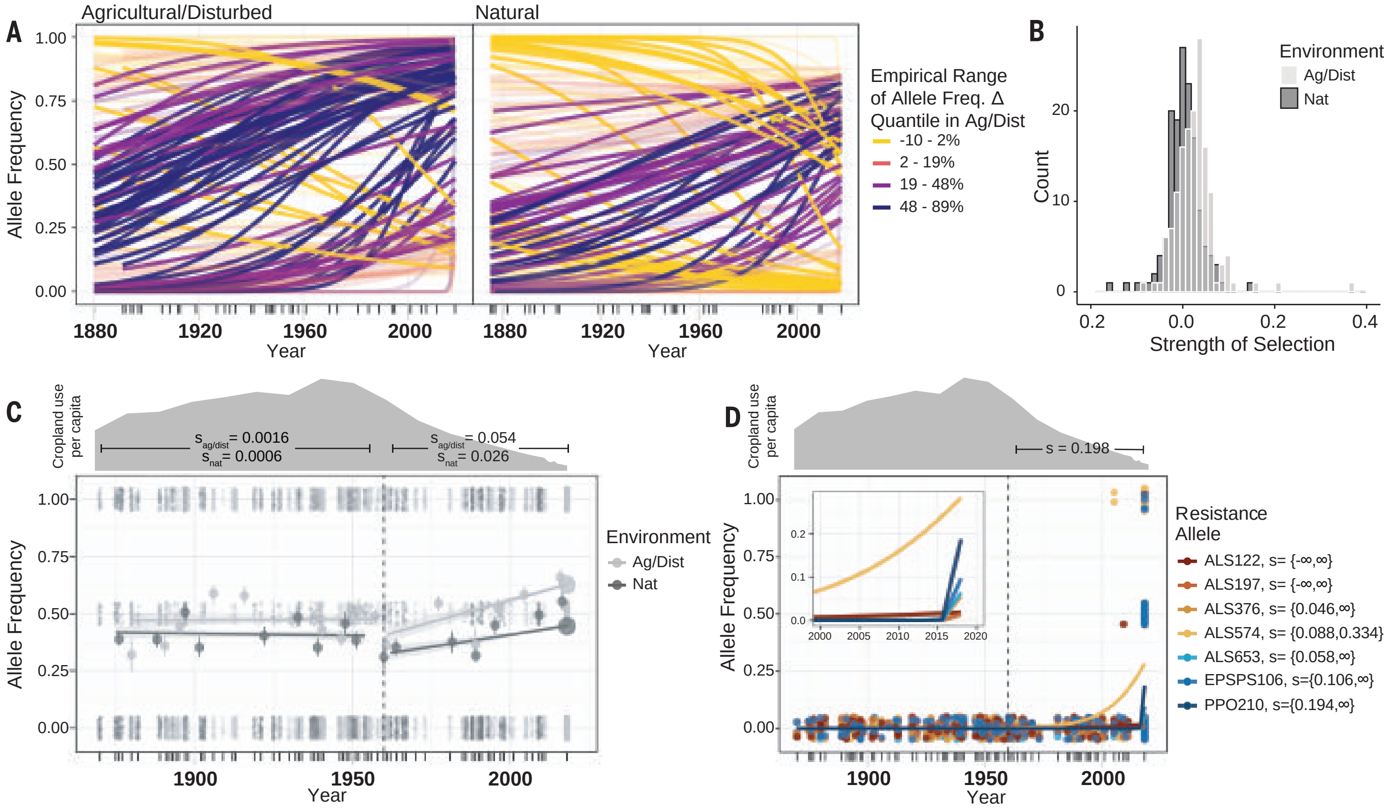


Fig. 3. Genomic signatures of agricultural adaptation through time. (**A**) Agricultural allele frequency trajectories for each of the 154 focal SNPs in agricultural and disturbed habitats (left) and in natural habitats (right). Trajectories are colored by the empirical range of the allele frequency change quantile in agricultural and disturbed habitats. Transparent lines indicate those with nonsignificant evidence of selection at α = 0.05 after FDR = 10% correction. (**B**) The distribution of selective strengths on agricultural alleles in natural (dark gray) and agricultural and disturbed (light gray) habitats between 1870 and 2018. (**C**) Environment-specific agricultural allele frequency trajectories before and after the start of agricultural intensification in 1960 (vertical dashed line). Large circles represent moving averages (over both loci and individuals) of allele frequencies, whereas dots represent raw genotype data for each locus and sample from which the allele frequency trajectory is estimated. Cropland use per capita in North America data are from (*1*), reflecting the intensity of agricultural practices. (**D**) The trajectory of alleles at known herbicide-resistance loci through time, fit by logistic regression for each of the biallelic resistance alleles present in our contemporary data (excluding EPSPSamp with its complex allelic structure). Dots represent genotypes for each historical and contemporary sample at each herbicide-resistance locus. The 95% CIs of the maximum likelihood estimates of selection between 1960 and 2018 are provided in the legend for each resistance allele.

over the past ~150 years include SNPs that map near *PPO*, *ACO1*, *CCB2*, *WRKY13*, *BPL3*, and *ATPD* (table S4). We find that both the total frequency change of agriculture-associated alleles and the estimated strength of selection in agricultural and disturbed environments are positively correlated with the extent of contemporary linkage disequilibrium around these loci (the number of SNPs within 1 Mb with r^2 > 0.25) (frequency change: F = 5.16, P = 0.024, r = 0.12; strength of selection: F = 3.99, P = 0.048, r = 0.058; fig. S6), consistent with theoretical expectations for the genomic signatures of recent positive selection (*43*, *44*).

We next investigated how well the trajectory of modern agricultural alleles reflects the rise of industrialized agricultural regimes over the past century. When we split our samples into those that predate versus those that come after the intensification of agriculture during the Green Revolution, we find that the increase in frequency of agricultural alleles was negligible in agricultural and disturbed environments before the 1960s (predicted 1870-to-1960 change = 0.005). By contrast, change subsequent to 1960 nearly completely accounts for the observed rise in frequency of modern agricultural alleles (predicted 1960-to-2018 change = 0.219 versus total 1870-to-2018 change = 0.221) (Fig. 3C). Corresponding estimates of selection by logistic regression using only data from before 1960 show no evidence of selection on these loci in disturbed and agricultural habitats [$\tilde{s} = 0.0008$, null interquantile range = (−0.0044, 0.0020)] or in natural habitats [$\tilde{s} = 0.0006$, null interquantile range = (−0.004, 0.004)]. However, samples collected after 1960 reflect a marked shift in selection—a collective $\tilde{s} = 0.054$ in disturbed and agricultural environments and a collective $\tilde{s} = 0.028$ in natural environments [agricultural and disturbed null interquantile range = (0.0064, 0.0020); natural null interquantile range = (−0.0056, 0.0054)] (Fig. 3C and fig. S8). Together, these results suggest that although most contemporary agricultural alleles were present in historical populations, these alleles only became associated with agricultural and human-managed sites over the past century, on time scales and at rates consistent with the rapid uptake and intensification of agrochemicals, controlled irrigation, and mechanization in agriculture.

The historical trajectory of known herbicide-resistance alleles epitomizes extreme selection over the past 50 years (Fig. 3D). Five of seven known biallelic herbicide-resistance alleles present in our contemporary, paired-environment collections are absent from our historical samples, consistent with the suggested importance of resistance adaptation from de novo mutation (*13*, *45*) and a particularly recent increase in their frequency. Only 3 of 108 historical samples show variation for

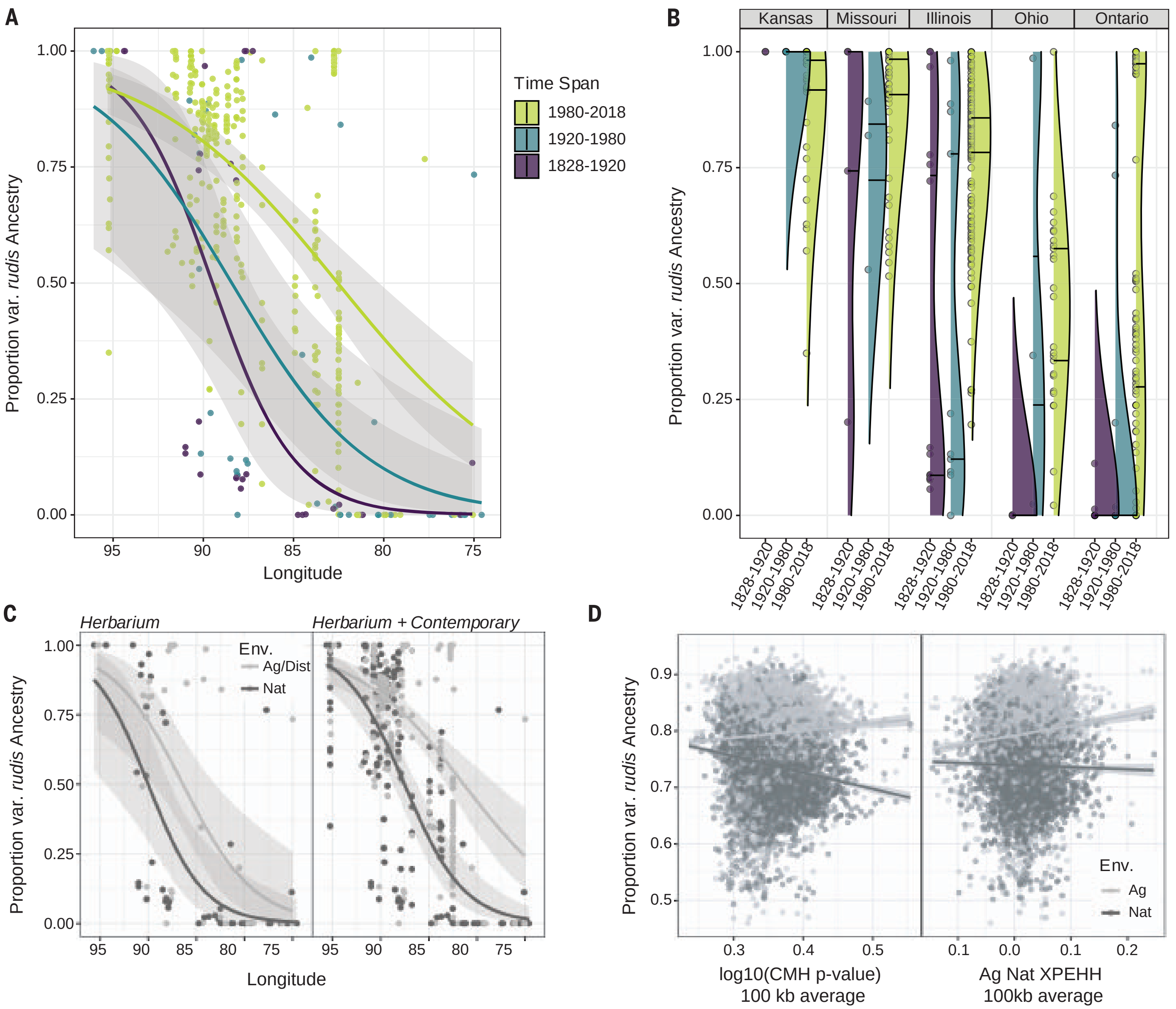


Fig. 4. Temporal shifts in the distribution of var. *rudis* ancestry have facilitated polygenic agricultural adaptation. (**A**) Longitudinal clines in var. *rudis* ancestry over three time spans, illustrating the expansion of var. *rudis* ancestry eastward over the past two centuries. In (A) to (C), dots represent individual-level ancestry estimates. (**B**) The distribution of individual-level var. *rudis* ancestry by state and through time, illustrating state-specific changes in ancestry. Horizonal lines within each distribution represent first and third quartiles of ancestry. (**C**) Increasing sorting of individual-level var. *rudis* ancestry into agricultural environments (Env.) on contemporary time scales. (**D**) Environment-specific metrics of selection [CMH *P* value and cross-population extended haplotype homozygosity (XPEHH)] across the genome in 100-kb windows positively correlate with var. *rudis* ancestry in agricultural but not natural habitats (XPEHH by environment: $F = 10.97$, $P = 9.3 \times 10^{-4}$; CMH by environment: $F = 108.51$, $P < 10^{-16}$).

herbicide resistance—two samples homozygous for resistance at ALS574 and one heterozygous for resistance at ALS122—all of which were sampled after the onset of herbicide applications in the 1960s (Fig. 3D). Resolving the very low historical and much higher contemporary frequencies of resistance, we estimate that since the approximate onset of herbicide use in 1960, these seven resistance alleles have collectively experienced a selective strength of $\tilde{s} = 0.198$ (logistic Z value = 2.11, $P = 0.035$) per year across environment types. Maximum likelihood–based estimates of selective strengths for each resistance allele are significant for five of the seven and are strongest for PPO210 ($s > 0.194$), EPSPS106 ($s > 0.106$), and ALS574 ($s > 0.088$) (Fig. 3D and table S3).

Concurrent temporal shifts in ancestry underlie agricultural adaptation

Finally, we explored whether historical demographic change over the past two centuries has played a role in agricultural adaptation. Early taxonomy described two different *A. tuberculatus* varieties as separate species, with few distinguishing characteristics [seed dehiscence and tepal length (*16*)]. Sauer's 1955 revision of the genus, which used herbarium specimens to gauge the distribution and migration of congeners over the past two centuries (*46*), led him to describe an expansion of the southwestern var. *rudis* type [at the time, *A. tamariscinus* (Sauer)] northeastward into the territory of var. *tuberculatus* [*A. tuberculatus* (Sauer)] sometime between 1856 to 1905 and 1906 to 1955. Our sequencing of >100 herbarium

samples dating back to 1828, combined with 349 contemporary sequences (*25*, *47*), allowed us to directly observe the change in the distribution of these two ancestral types, adding further temporal resolution to Sauer's morphological observations of the species' range shifts, and to assess the role of agriculturally adaptive standing genetic variation across varieties.

Range-wide, we see clear shifts in the distribution of var. *rudis* ancestry based on fastSTRUCTURE (*48*) inference at K (number of subpopulations) = 2 (fig. S9) across three time spans—1830 to 1920, 1920 to 1980, and 1980 to 2018 (time span: $F = 5.47$, $P = 0.0045$)—and particularly so in the east (time span × longitude: $F = 5.49$, $P = 0.0045$), consistent with a recent expansion of var. *rudis* ancestry (Fig. 4A). Furthermore, we see strong state- and province-specific shifts in ancestry through time in our historical sequences (time span by state interaction: $F = 4.22$, $P = 7 \times 10^{-5}$), highlighting not only the shift of var. *rudis* eastward (with increases through time in Ontario, Ohio, Illinois, and Missouri) but also the recent introduction of var. *tuberculatus* ancestry into the most western part of the range in Kansas (Fig. 4B). *A. tuberculatus* demography thus appears to have been markedly influenced by human-mediated landscape change over the past two centuries, consistent with the massive recent expansion of effective population size that we had previously inferred from contemporary samples over this same time frame (*45*). That this shift has been most notable over the past 40 years is further consistent with the time scale of agricultural intensification, shifts toward conservation tillage, and rampant herbicide-resistance evolution within the species (*19*, *45*, *49*, *50*), which suggests that selection may facilitate the colonization of var. *rudis* ancestry outside its historical range. Along these lines, we find that this contemporary range expansion has facilitated the sorting of var. *rudis* ancestry across environments (a longitude by time span by environment interaction: $F = 5.13$, $P = 4 \times 10^{-5}$; Fig. 4C), with increasing overrepresentation of var. *rudis* ancestry in agricultural and disturbed environments in the eastern portion of the range through time, as has been previously suggested (*25*).

To investigate whether agricultural adaptation has drawn disproportionately from var. *rudis* ancestry, we reconstructed fine-scale ancestry across the genome. On the basis of analyses in 100-kb windows, we find a least squares mean of 5.5% [95% confidence interval (CI) = (5.0, 5.9%)] more var. *rudis* ancestry genome-wide in agricultural environments compared with the adjacent natural habitat (fig. S10). The environment-specific proportion of var. *rudis* ancestry is not only positively correlated with recombination rate ($F = 18.85$, $P = 1.4 \times 10^{-5}$, $r = 0.056$) and gene density ($F = 8.53$, $P = 0.004$, $r = 0.050$) but also with SNP and haplotype-based evidence of environment-specific selection. Agricultural but not natural populations have an excess of cross-population haplotype homozygosity (agricultural versus natural XPEHH) and within-pair environmental differentiation (CMH P value) in genomic regions of high var. *rudis* ancestry (XPEHH by environment: $F = 10.97$, $P = 9.3 \times 10^{-4}$; CMH by environment: $F = 108.51$, $P < 10^{-16}$; Fig. 4D), which implies that ancestry composition genome-wide in large part determines the extent of polygenic agricultural adaptation. These findings suggest that the expansion of var. *rudis* ancestry across the range, particularly in the past 40 years, has facilitated waterhemp's success in agricultural habitats by providing access to preadapted, standing genetic variation.

Discussion

Agricultural adaptation in *A. tuberculatus*, a native plant in North America, has occurred over extremely short time scales, facilitated by range shifts in response to the agriculturalization of its native habitat. The human-mediated expansion of the southwestern lineage of the species northeastward since the latter half of the 20th century has introduced new genetic variation across the range, on which selection in agricultural settings could act. Negligible genome-wide differentiation across habitats in this species refutes the idea of agricultural populations existing as separate to natural ecosystems (*51*). Despite substantial gene flow, the prevalence of agricultural alleles has increased rapidly since the intensification of agriculture over the past 60 years—in agricultural environments by a selection coefficient of nearly 6% per year and even in natural habitats by >2% per year. The selective intensity of industrial agriculture is on par with the intensities that *Arabidopsis* populations in extreme hot and dry environments are predicted to face by 2070 under the high-emissions scenario of climate change (*52*). The effects of agricultural herbicides are even more extreme—range-wide, evolved resistance mutations have experienced selective strengths of 20% on average per year since herbicides were first introduced—permeating even into natural habitats.

Although modern, industrial agriculture imposes strengths of selection rarely observed in the wild, *A. tuberculatus* has in turn escalated the weed management-evolution arms race through a multitude of interdependent mechanisms: range expansion, polygenic adaptation from standing genetic variation, and large-effect herbicide-resistance mutations. Together, these results highlight that anthropogenic change not only leads to the formation of new habitats but also provides an opportunity for range expansion that may facilitate and interact with local adaptation, reshaping genetic variation for fitness within native species.

REFERENCES AND NOTES

1. K. Klein Goldewijk, A. Beusen, J. Doelman, E. Stehfest, *Earth Syst. Sci. Data* **9**, 927–953 (2017).
2. G. F. Sassenrath *et al.*, *Renew. Agric. Food Syst.* **23**, 285–295 (2008).
3. T. E. Crews, W. Carton, L. Olsson, *Glob. Sustain.* **1**, E11 (2018).
4. E. Malaj, L. Freistadt, C. A. Morrissey, *Front. Environ. Sci.* **8**, 556452 (2020).
5. J. Fernandez-Cornejo, "Pesticide Use in U.S. Agriculture: 21 Selected Crops, 1960-2008" (USDA-ERS Economic Information Bulletin, Number 124, 2014); http://dx.doi.org/10.2139/ssrn.2502986.
6. N. E. Borlaug, *Science* **219**, 689–693 (1983).
7. H. J. Beckie *et al.*, *Can. J. Plant Sci.* **86**, 1243–1264 (2006).
8. C. Mann, *Science* **277**, 1038–1043 (1997).
9. P. L. Pingali, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 12302–12308 (2012).
10. P. Pellegrini, R. J. Fernández, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 2335–2340 (2018).
11. J. Mallet, *Trends Ecol. Evol.* **4**, 336–340 (1989).
12. C. Délye, M. Jasieniuk, V. Le Corre, *Trends Genet.* **29**, 649–658 (2013).
13. N. J. Hawkins, C. Bass, A. Dixon, P. Neve, *Biol. Rev.* **94**, 135–155 (2018).
14. F. Gould, Z. S. Brown, J. Kuzma, *Science* **360**, 728–732 (2018).
15. F. Zabel *et al.*, *Nat. Commun.* **10**, 2844 (2019).
16. J. D. Sauer, *Madroño* **13**, 5–46 (1955).
17. K. E. Waselkov, K. M. Olsen, *Am. J. Bot.* **101**, 1726–1736 (2014).
18. M. Costea, S. E. Weaver, F. J. Tardif, *Can. J. Plant Sci.* **85**, 507–522 (2005).
19. P. J. Tranel, *Pest Manag. Sci.* **77**, 43–54 (2021).
20. J. Hermisson, P. S. Pennings, *Genetics* **169**, 2335–2352 (2005).
21. R. D. H. Barrett, D. Schluter, *Trends Ecol. Evol.* **23**, 38–44 (2008).
22. M. Alleaume-Benharira, I. R. Pen, O. Ronce, *J. Evol. Biol.* **19**, 203–215 (2006).
23. R. I. Colautti, S. C. H. Barrett, *Science* **342**, 364–366 (2013).
24. M. Szűcs *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 13501–13506 (2017).
25. J. M. Kreiner, A. Caballero, S. I. Wright, J. R. Stinchcombe, *Evolution* **76**, 70–85 (2022).
26. F. E. Dayan, S. O. Duke, in *Hayes' Handbook of Pesticide Toxicology*, R. Krieger, Ed. (Academic Press, 2010), pp. 1733–1751.
27. J. G. Moraes *et al.*, *Agronomy* **11**, 754 (2021).
28. W. Moeder, O. Del Pozo, D. A. Navarre, G. B. Martin, D. F. Klessig, *Plant Mol. Biol.* **63**, 273–287 (2007).
29. P. A. Ribone, M. Capella, R. L. Chan, *J. Exp. Bot.* **66**, 5929–5943 (2015).
30. J. V. Cabello, R. L. Chan, *Plant Biotechnol. J.* **10**, 815–825 (2012).
31. S. Guénin *et al.*, *J. Exp. Bot.* **68**, 1083–1095 (2017).
32. S. Zhou *et al.*, *PLOS Genet.* **12**, e1006255 (2016).
33. C. Dai, Y. Lee, I. C. Lee, H. G. Nam, J. M. Kwak, *Front. Plant Sci.* **9**, 803 (2018).
34. H. Guo, H. Yang, T. C. Mockler, C. Lin, *Science* **279**, 1360–1363 (1998).
35. T. Mockler *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 2140–2145 (2003).
36. W. Wang *et al.*, *Plant Cell* **30**, 1989–2005 (2018).
37. J. Li, Y. Li, S. Chen, L. An, *J. Exp. Bot.* **61**, 4221–4230 (2010).
38. F. E. Dayan *et al.*, *Biochim. Biophys. Acta* **1804**, 1548–1556 (2010).
39. H. M. Cockerton *et al.*, *Front. Plant Sci.* **12**, 10.3389/fpls.2021.651381 (2021).
40. M. M. Vila-Aiub, *Plants* **8**, 469 (2019).
41. Materials and methods are available as supplementary materials.
42. N. H. Barton, *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2122147119 (2022).
43. M. Przeworski, *Genetics* **160**, 1179–1189 (2002).
44. Y. Kim, R. Nielsen, *Genetics* **167**, 1513–1524 (2004).
45. J. M. Kreiner *et al.*, *eLife* **11**, e70242 (2022).
46. J. Sauer, *Evolution* **11**, 11–31 (1957).
47. J. M. Kreiner *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 21076–21084 (2019).
48. A. Raj, M. Stephens, J. K. Pritchard, *Genetics* **197**, 573–589 (2014).
49. M. J. Foes, L. Liu, P. J. Tranel, L. M. Wax, E. W. Stoller, *Weed Sci.* **46**, 514–520 (1998).
50. P. J. Tranel, C. W. Riggins, M. S. Bell, A. G. Hager, *J. Agric. Food Chem.* **59**, 5808–5812 (2011).
51. Q. C. B. Cronk, J. L. Fuller, *Plant Invaders: The Threat to Natural Ecosystems* (Routledge, 2014).

52. M. Exposito-Alonso *et al*., *Nature* **573**, 126–129 (2019).
53. J. Kreiner, jkreinz/TemporalAdaptation: Oct92022_TemporalAdaptationCode, version 1.0.0, Zenodo (2022); https://doi.org/10.5281/zenodo.7178764.

ACKNOWLEDGMENTS

We appreciate the pivotal contributions of numerous herbaria toward this research, especially the help of E. Knox at the Indiana University Herbarium, J. L. Minnaert-Grote at the University of Illinois INHS Herbarium, T. Mesfin at the University of Ohio Herbarium, A. Reznicek at the University of Michigan Herbarium, J. Solomon at the Missouri Botanical Gardens, C. Morse at the McGregor Herbarium at the University of Kansas, T. Smith and S. Wang at Agricultural and Agrifood Canada, and D. Metsger and T. Dickinson at the Royal Ontario Museum. We thank M. Whitlock and T. Booker (University of British Columbia), A. Agrawal and T. Kent (University of Toronto), and A. Macpherson (Simon Fraser University) for input on the work; C. Lanz and R. Schwab (Max Planck Institute) for coordinating the sequencing of herbarium samples; E. Reiter (University of Leipzig) for scheduling and coordinating logistics for clean room facility work; and P. Lang, S. Kersten, and H. Budde (Max Planck Institute) for advice on molecular protocols troubleshooting. **Funding:** J.M.K. was supported by the Biodiversity Research Institute at the University of British Columbia and a Killam Fellowship. S.I.W. was supported by the NSERC (RGPIN-2020-05850) and a Canada research chair. J.R.S. was supported by the NSERC (RGPIN-2022-04366). S.P.O. was supported by NSERC RGPIN-2022-03726. S.M.L., H.A.B., and D.W. were supported by the Max Planck Society. **Author contributions:** J.M.K., J.R.S., and S.I.W. conceptualized the paired sampling design; J.M.K., H.A.B., D.W., J.R.S., and S.I.W. conceptualized the use of herbarium data; J.M.K. performed contemporary collections and curated the herbarium samples; S.M.L. and H.A.B. conceptualized and designed the molecular work with herbarium specimens; S.M.L. coordinated the clean room facility work; J.M.K. and S.M.L. performed DNA extraction and library preparations of herbarium tissue; and S.M.L. oversaw the sequencing of herbarium specimens. J.M.K. performed analyses with input from S.P.O., S.I.W., and J.R.S. S.P.O. wrote the migration-selection balance and maximum likelihood models. J.M.K. wrote and revised the paper with inputs from all authors. **Competing interests:** D.W. holds equity in Computomics, which advises breeders. D.W. consults for KWS SE, a plant breeder and seed producer. The authors declare no other competing interests. **Data and materials availability:** All new sequence data have been archived at the NCBI Sequence Read Archive (BioProject ID PRJNA878842). Scripts and accompanying metadata have been archived on Github (www.github.com/jkreinz/TemporalAdaptation) and on Zenodo (*53*). **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abo7293
Materials and Methods
Figs. S1 to S13
Tables S1 to S4
References (*54*–*68*)
Data S1 and S2

Submitted 25 February 2022; resubmitted 24 June 2022
Accepted 17 October 2022
10.1126/science.abo7293

ORGANIC CHEMISTRY

Catalytic asymmetric C–H insertion reactions of vinyl carbocations

Sepand K. Nistanaki[1], Chloe G. Williams[1], Benjamin Wigman[2], Jonathan J. Wong[2], Brittany C. Haas[3], Stasik Popov[2], Jacob Werth[3], Matthew S. Sigman[3]*, K. N. Houk[2]*, Hosea M. Nelson[1]*

From the preparation of pharmaceuticals to enzymatic construction of natural products, carbocations are central to molecular synthesis. Although these reactive intermediates are engaged in stereoselective processes in nature, exerting enantiocontrol over carbocations with synthetic catalysts remains challenging. Many resonance-stabilized tricoordinated carbocations, such as iminium and oxocarbenium ions, have been applied in catalytic enantioselective reactions. However, their dicoordinated counterparts (aryl and vinyl carbocations) have not, despite their emerging utility in chemical synthesis. We report the discovery of a highly enantioselective vinyl carbocation carbon–hydrogen (C–H) insertion reaction enabled by imidodiphosphorimidate organocatalysts. Active site confinement featured in this catalyst class not only enables effective enantiocontrol but also expands the scope of vinyl cation C–H insertion chemistry, which broadens the utility of this transition metal–free C(sp^3)–H functionalization platform.

Since the late 19th century, carbocations (cationic molecules that feature an even number of electrons with a positive charge residing on one or more carbon atoms) have fascinated organic chemists because of their fundamental properties and potent modes of reactivity (*1*). From agostic bonding, which is critical to our understanding of chemical structure (*2*, *3*), to biosynthesis, in which carbocations facilitate the assembly of stereochemically complex natural products (*4*, *5*), these reactive intermediates have played an essential role in our understanding of chemistry. Despite the relevance of carbocations to many modern synthetic processes, controlling the regio-, chemo-, and stereoselectivity of their reactions has proven challenging (*6*). This is especially true in asymmetric catalysis, in which high-energy, prochiral cationic intermediates must proceed through energetically differentiated diastereomeric transition states (TSs) to obtain synthetically useful enantiomeric excesses (ee). Achieving an energy difference ($\Delta\Delta G^{\ddagger}$) of >2.0 kcal/mol (which is required for >90% ee at 70°C) is a daunting task for such high-energy intermediates for which productive bond formation can proceed via relatively low energetic barriers (Fig. 1A). Moreover, the propensity for such intermediates to engage in unproductive side reactions further complicates the development of selective processes. This fundamental challenge is manifested in modern carbocation chemistry, in which high levels of selectivity are only routinely realized in the reactions of resonance- or heteroatom-stabilized carbocationic species (*6*–*10*). The asymmetric reactions of carbocations that are not stabilized by π-resonance remain rare (*11*).

Unlike resonance-stabilized tricoordinated carbocations, dicoordinated carbocations, such as aryl and vinyl cations, have thus far been excluded from the field of asymmetric catalysis (Fig. 1B). This is likely due to the lack of catalytic methods for their generation and their high reactivity once formed (*12*). Recent reports that these intermediates undergo facile insertion into unactivated sp^3 C–H bonds (*13*–*20*), which often creates stereogenic carbons, prompted us to investigate their application in asymmetric catalysis. In addition to the fundamental challenges associated with engaging a high-energy species in catalytic enantioselective reactions [in some cases, dicoordinated cations are computed to undergo almost barrierless insertion into alkane C–H bonds (*15*, *16*, *20*)], most catalytic C–H insertion reactions of dicoordinated carbocations have required the use of weakly coordinating anions (WCAs). These anions are notable in their weak basicity and nucleophilicity, which avoids deleterious E1 elimination or S_N1 nucleophilic substitution reactions (*21*). However, the few chiral WCAs that have been reported have not been successfully used in asymmetric catalysis (*22*, *23*) because the highly ordered TSs that are required for enantioselectivity almost always rely on directional coordinative interactions. Inspired by the exquisite stereocontrol over reactive carbocations manifested within a confined enzyme active site (*24*) and key synthetic studies on confinement effects in selective catalysis (*25*, *26*), we hypothesized that confinement of a dicoordinated carbocation intermediate could direct its ensuing reactivity with enantiocontrol. In this work, we report the discovery of a highly enantioselective C–H insertion reaction of vinyl carbocations enabled by imidodiphosphorimidate (IDPi) strong acids pioneered by List *et al.*

[1]Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125, USA.
[2]Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, CA 90095, USA.
[3]Department of Chemistry, University of Utah, Salt Lake City, UT 84112, USA.
***Corresponding author. Email: hosea@caltech.edu (H.M.N.); houk@chem.ucla.edu (K.N.H.); matt.sigman@utah.edu (M.S.S.)**

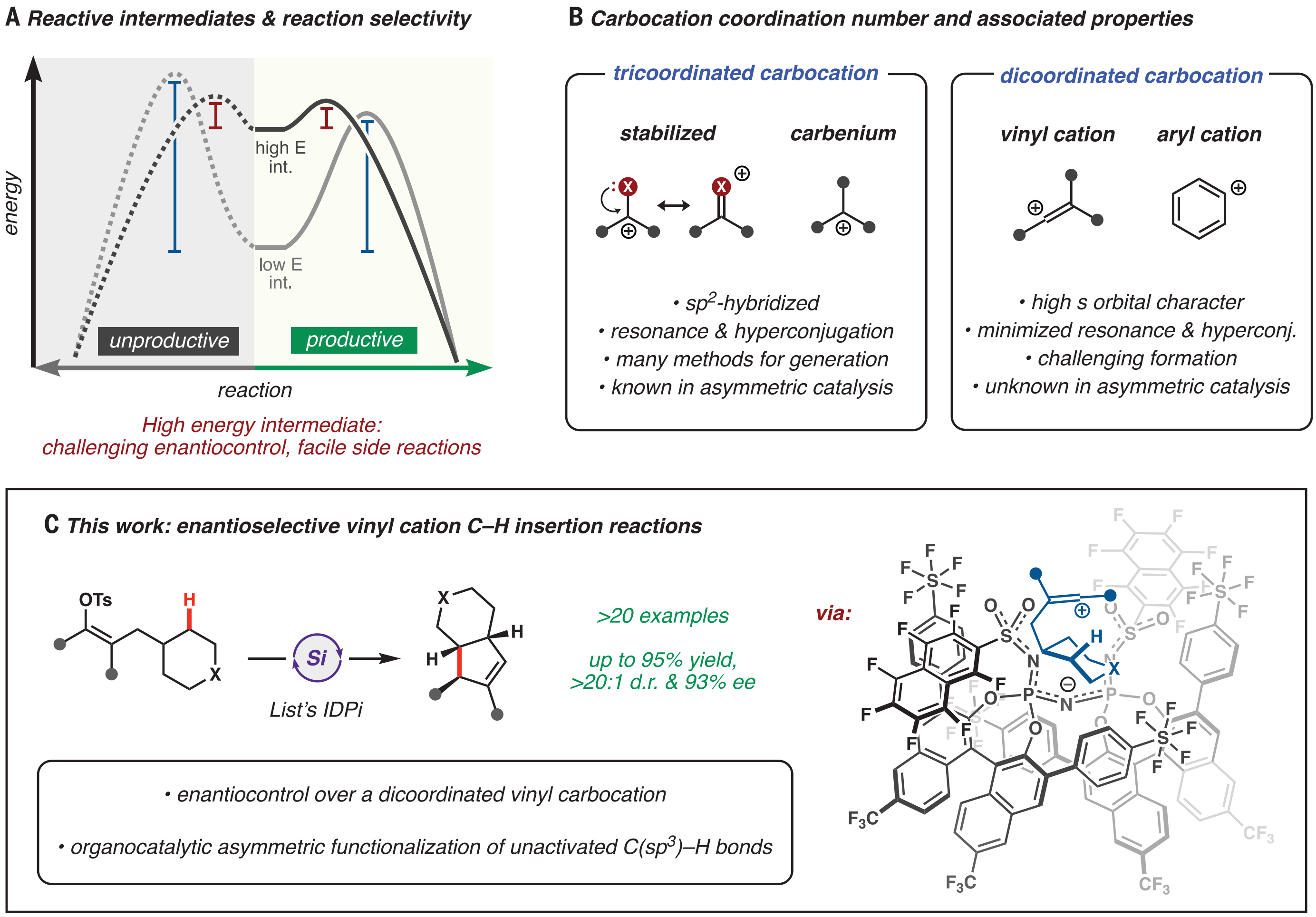


Fig. 1. Selective reactions of carbocations. (**A**) Challenges associated with high-energy intermediates in selective catalysis. (**B**) Properties associated with carbocation coordination number. (**C**) Discovery of enantioselective C–H insertion reactions of vinyl cations. high E int, high-energy intermediate; low E int, low-energy intermediate; OTs, *p*-toluenesulfonate; X, NSO_2CF_3 or CH_2.

(*27*, *28*), which results in an organocatalytic platform for stereoselective functionalization of unactivated C(sp^3)–H bonds (Fig. 1C).

Our initial investigations focused on the intramolecular desymmetrization of a piperidinyl fragment via C–H insertion to generate bicyclic structures, given the prevalence of nitrogen-containing polycyclic scaffolds in numerous bioactive natural products (*29*). After exploration of various catalyst structures (Table 1), we were able to catalyze the transformation of vinyl tosylate **1a** to the cyclized C–H insertion product **2a** with high ee using IDPi **8B** and stoichiometric allyltriisopropylsilane (entry 6). This C–H functionalization reaction forges three contiguous stereocenters with notable diastereoselectivity [>20:1 diastereomeric ratio (d.r.)] and reasonable conversion after 72 hours. We posit that the reaction is sluggish because of a high barrier for vinyl carbocation formation and therefore is directly dependent on the Lewis acidity of the silicon species (*12*). Thus, we hypothesized that the conversion could be improved by tuning the silyl group on the allyl silane, given that the silylated IDPi is likely the active catalyst (see below). We were inspired by Lambert's studies on the effects of steric bulk on silylium ion coordination, in which tris(trimethylsilyl)silylium [$Si(TMS)_3^+$] paired with a WCA generated a near-trivalent silylium cation, in contrast to trialkyl silylium cations that form tetravalent Si centers by coordination to solvent or counteranion (*30*). Gratifyingly, in our system, we observed a positive correlation between silane size and activity (entries 6 to 8 in Table 1), in which tris(triethylsilyl) allylsilane resulted in full conversion without any influence on enantioselectivity. We performed Gutmann-Beckett studies that supported the enhanced Lewis acidity of tris(triethylsilyl)silane [$Si(TES)_3$]–silylated IDPi relative to its trialkyl silyl IDPi counterpart, although other effects cannot be ruled out (see supplementary materials).

With this improved activity and encouraging enantioselectivity on our model substrate, we explored the scope of this reaction (Fig. 2). The transformation proved compatible with substitution at both of the aryl rings on the substrate, delivering insertion products in moderate to good yield with excellent enantioselectivity (up to 93% ee) and diastereoselectivity (>20:1 d.r.). High levels of enantioselectivity were maintained with ortho substitution near the insertion site (**2g** and **2i**), with electron-rich (**2f**, **2g**, and **2o**), electron-deficient (**2e**, **2n**, and **2p**), and heterocyclic (**2j** and **2k**) substrates also performing well. Moreover, functional groups labile in many transition metal–catalyzed processes (**2p** to **2s**) were compatible, which highlights this method's complementarity to transition metal–catalyzed C(sp^3)–H functionalization platforms. C–H insertion adjacent to a sterically congested all-carbon quaternary center (**2u**) occurred in good yield, albeit with moderate enantioselectivity. An additional highlight was the conversion of an all-hydrocarbon variant of the model substrate to bicyclo[4.3.0]nonenes in modest stereoselectivity (**2v**). The enantiopurity of many of the piperidine insertion products could be readily upgraded via a single recrystallization to afford enantiopure material

(>99% ee). Moreover, the *N*-triflyl protecting group could be removed with sodium bis(2-methoxyethoxy)aluminum hydride (Red-Al) to afford free amine **2w** (*31*). Heating insertion product **2a** with acid isomerized the olefin to the thermodynamically more favored tetrasubstituted olefin isomer (**2x**), without loss of enantioenrichment. WCA catalysts that were previously used in our laboratory in vinyl cation C–H insertion reactions resulted in low yields (<10%) or intractable product mixtures because of the formation of multiple isomers, oligomeric products, and other unidentifiable side products (see supplementary materials). These findings, when compared with previous examples of catalytic vinyl cation C–H insertion reactions, highlight a compelling advance in substrate compatibility and catalyst control.

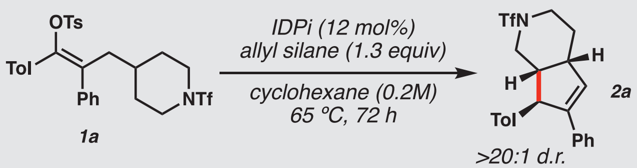
Confirming the intermediacy of the vinyl cation

Although catalytic vinyl cation formation via sulfonate abstraction has been previously reported (*15*, *16*), we sought to further support its intermediacy in this system. To this end, we found that both the *Z*- and *E*-vinyl tosylate isomers afforded the same insertion product with identical enantioselectivities, which supports the common intermediacy of a linear vinyl carbocation (Fig. 3A). Moreover, we prepared tosylate **9** (which cannot undergo intramolecular C–H insertion) and found that allene **10** is produced as the major product under the reaction conditions, which likely arises from a Wagner-Meerwein rearrangement—a known reactivity pathway of vinyl carbocations (*12*).

In the context of enantioselective insertion, the allyl silane did not influence the selectivity of the reaction, despite its profound influence on rate. This result not only supports the proposed role of the Si-center in the ionization of the substrate but also suggests that after ionization, a discrete vinyl cation–IDPi ion pair is formed that undergoes the enantiodetermining C–H insertion step. Supporting this hypothesis, deuteration of the substrate's insertion sites resulted in an appreciable increase in % ee across two structurally distinct catalysts (Fig. 3B), which corresponds to up to a 0.12 ± 0.03 kcal/mol difference in $\Delta\Delta G^{\ddagger}$ (*32*). Moreover, no indication of deuterium incorporation adjacent to the CD_2 methylene was observed, which would be expected if a stepwise hydride rebound mechanism was operative (*15*) given that hydride shifts would result in H/D scrambling (Fig. 3C). Additionally, substrate **1u** cleanly forms insertion product **2u** in high yield, with no sign of rearrangement. Although these observations support either a concerted insertion or a rapid rebound mechanism, in which a putative 2° carbocation is trapped by olefin attack before facile 1,2-shifts can occur, we recognize that the mechanism likely resides on a continuum between these two extreme scenarios (see below). On the basis of these results, a proposed mechanism (Fig. 3D) initiates with protodesilylation of the allylsilane to generate the active Lewis acid **11**, which is consistent with prior work by List *et al.* (*27*, *33*). Subsequent ionization of the vinyl tosylate by **11** generates the reactive vinyl cation species (**12**), which we propose is confined within the IDPi anion as a contact ion pair. An enantiodetermining C–H insertion step followed by kinetic deprotonation delivers insertion product **2** and regenerates IDPi.

Modeling the TS

To probe the specific interactions responsible for the experimentally observed stereoselectivity, density functional theory (DFT) calculations at the ωB97X-D/6-311++G(d,p), CPCM(cyclohexane)//B3LYP-D3/6-31G(d,p), CPCM(cyclohexane) level were carried out. Two diastereomeric TSs that used substrate **1a** and catalyst **8B** were computed and found to correspond with a concerted, asynchronous C–H insertion of the vinyl cation with an energy difference ($\Delta\Delta G^{\ddagger}$) of 2.1 kcal/mol, which favored the experimentally observed enantiomer (Fig. 3E). In the minor TS, the sterically encumbering piperidine is situated in much closer proximity to the congested *p*-SF_5 substituents of the catalyst scaffold. These repulsive interactions obstruct the optimal positioning of the substrate within the anionic pocket of the catalyst, which ultimately weakens the intermolecular interactions between the associated ions. The difference in electrostatic stabilization between the two TSs was calculated to be 0.6 kcal/mol (*34*) with an additional 1.4 kcal/mol of stabilization in the favored enantiomeric TS arising from dispersive interactions, which together correspond to enhanced binding of the C–H insertion TS by the IDPi anion if the preferred enantiomeric conformation is adopted. Careful inspection of the TS structure reveals bending of the vinyl cation, placing the tolyl group away from the bulky



Entry	IDPi	X	Silane	Conv.	Yield	%ee
1	3A	H	allyl TIPS	70%	56%	52%
2	4B	CF_3	allyl TIPS	89%	79%	85%
3	5B	CF_3	allyl TIPS	95%	72%	60%
4	6B	CF_3	allyl TIPS	13%	11%	84%
5	7B	CF_3	allyl TIPS	93%	84%	85%
6	8B	CF_3	allyl TIPS	81%	72%	91%
7	8B	CF_3	allyl TMS	38%	34%	89%
8	8B	CF_3	allyl $Si(TES)_3$	full	91% (81%)*	91%
9	8B	CF_3	none	0%	0%	—

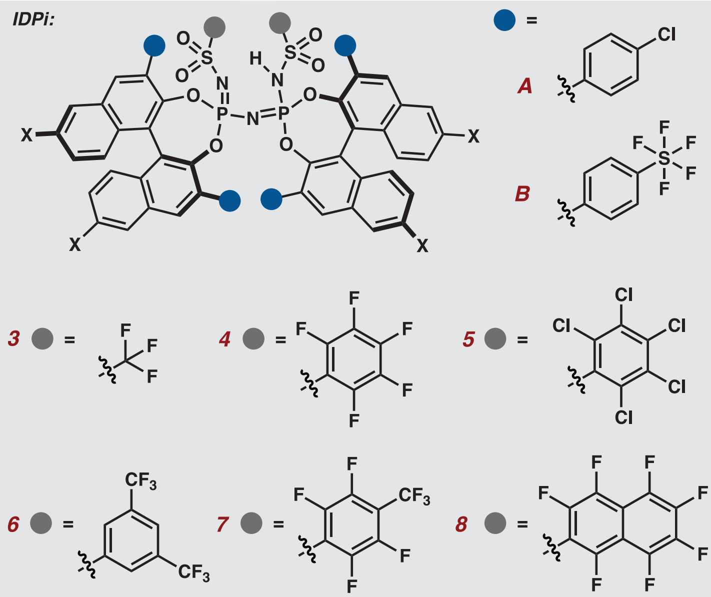


Table 1. Reaction optimization. Yields determined by [1]H NMR with an internal standard. *Isolated yield on 0.1 mmol scale. conv., conversion of starting material; TIPS, triisopropylsilane; TMS, trimethylsilane.

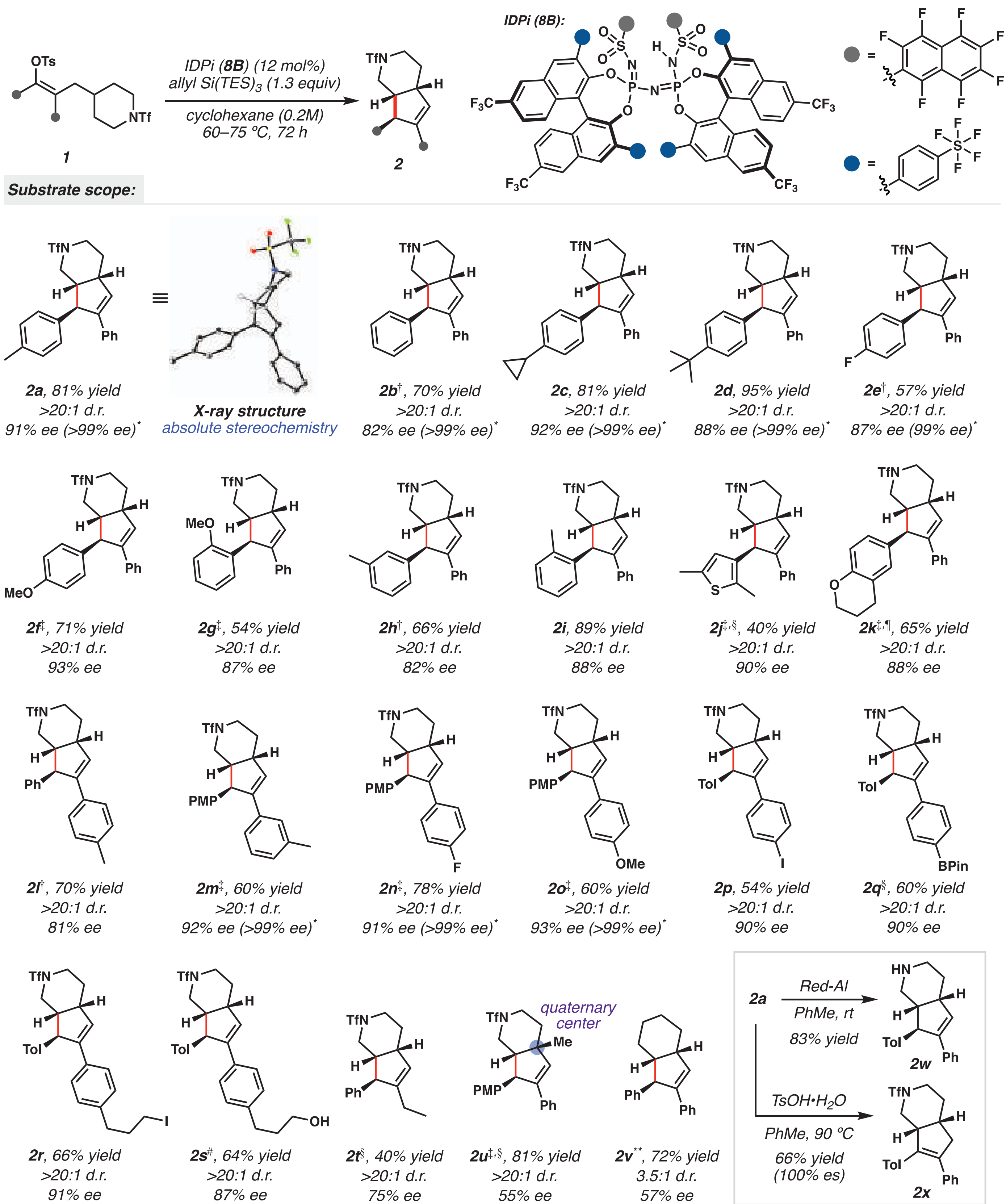


Fig. 2. Substrate scope. All reported yields are isolated yields of purified product. *After single recrystallization. †96 hours at 75°C. ‡0.1M. §70°C. ¶60°C. #2.3 equivalents of silane used because of alcohol silylation; the crude reaction was then stirred with tetrabutylammonium fluoride. **0.025M using **3A**. BPin, pinacol boronic ester; es, enantiospecificity; Me, methyl; Ph, phenyl; PhMe, toluene; PMP, *p*-methoxyphenyl; rt, room temperature; Tol, *p*-tolyl; TsOH, *p*-toluenesulfonic acid.

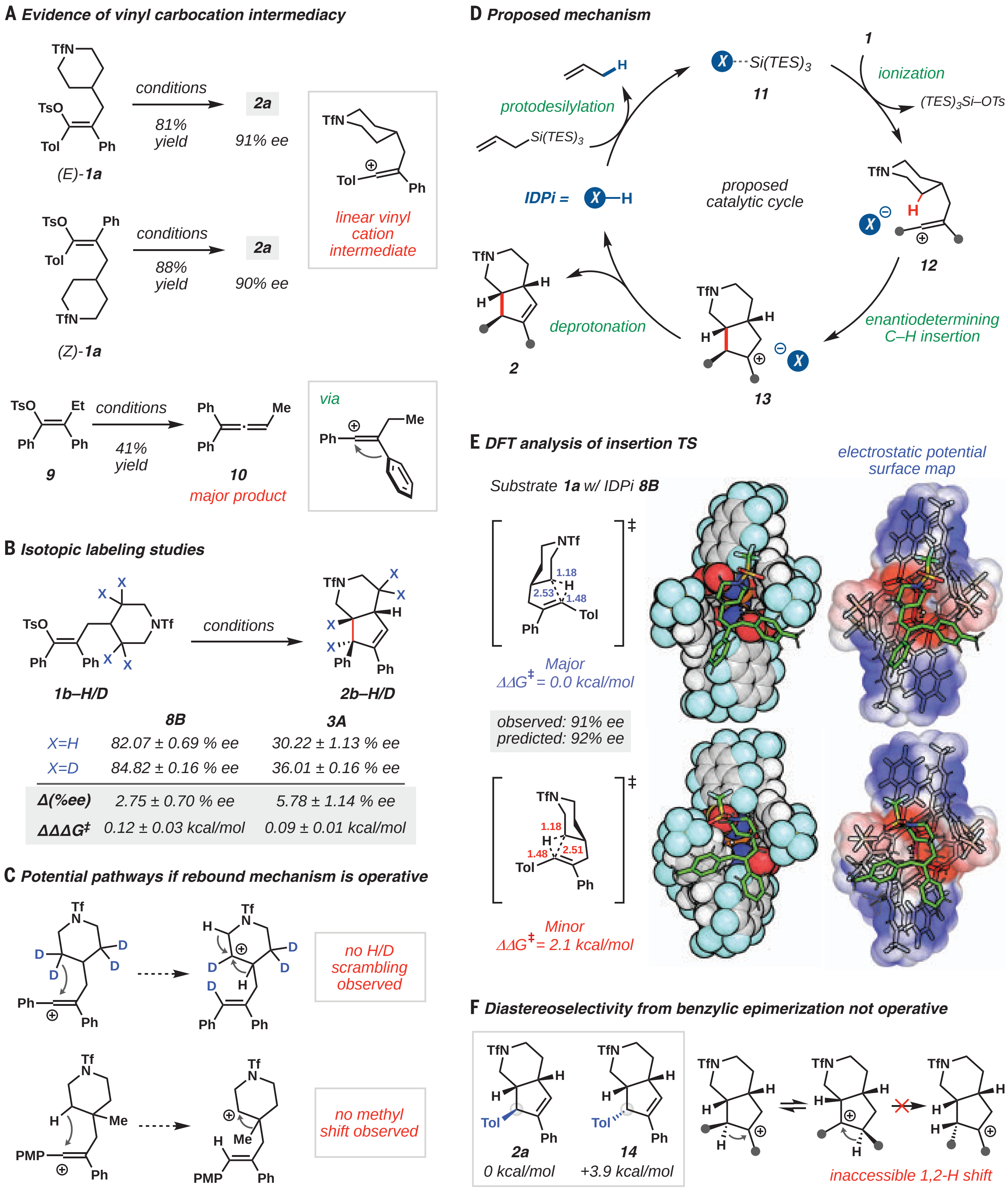


Fig. 3. Mechanistic studies. (**A**) Evidence for vinyl carbocation intermediacy. (**B**) Effect of isotopic labeling on enantioselectivity. Listed values are the average of triplicate runs. (**C**) Rearrangement products from a rebound mechanism that were not observed. (**D**) Proposed mechanism. (**E**) DFT-calculated diastereomeric TS structures with corresponding bond lengths (Å), electrostatic potential surfaces, and free-energy differences between major and minor TSs. Substrate atoms are rendered as a stick model, and the catalyst is rendered as a space-filling model. Electrostatic potential areas are colored red to indicate a more-negative potential and blue to indicate a more-positive potential. (**F**) Benzylic stereocenter epimerization is not likely. Et, ethyl.

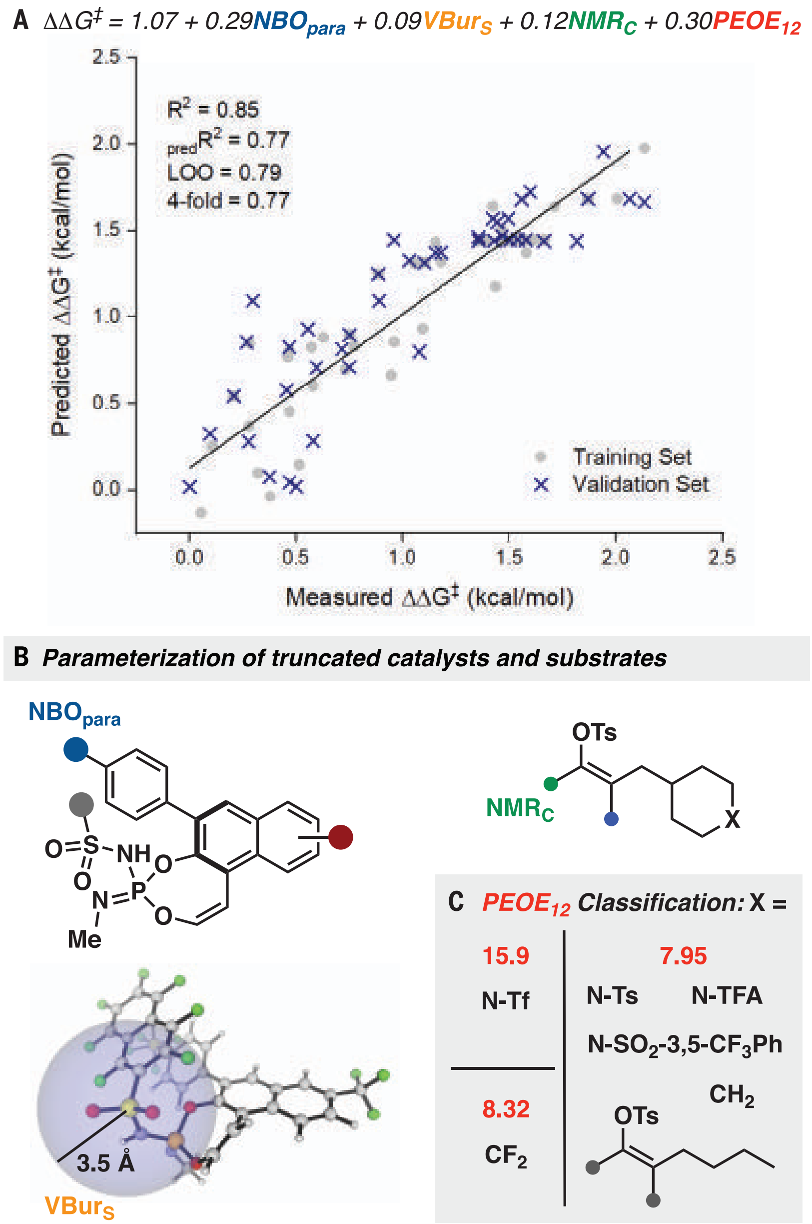


Fig. 4. Statistical modeling. (**A**) Multivariate linear regression model with a pseudorandom 50:50 partitioning of the 91 data points into training set:validation set. (**B**) Visual representation of the molecular descriptors used in the model. (**C**) Substrate classification of X identity on the basis of $PEOE_{12}$.

N–Tf piperidine ring (Tf, trifluoromethylsulfonyl), which is consistent with an anti relative orientation seen in the experimentally obtained diastereomer. Although product **2a** was computed to be energetically favored over its benzylic epimer (**14**; Fig. 3F), thermodynamically driven equilibration after C–H insertion is likely not the driver of diastereoselectivity. 1,2-hydride shifts are known to be stereospecific (*35*), which rules out benzylic hydride exchange as a potential mechanism for epimerization of the benzylic stereocenter (Fig. 3F), and hydride shifts from the ring fusion carbon are discounted given the lack of H/D scrambling observed in deuterated substrate **1b**–$\mathbf{D_4}$ (Fig. 3C). Taken together, these results are consistent with a 1,1-insertion step that sets all three stereocenters concurrently with high stereocontrol (see supplementary materials for further discussion).

To further explore factors governing how catalyst and substrate structure affected the enantioselective outcomes, we investigated the reaction holistically by using statistical modeling to analyze the diverse reaction outputs recorded within the optimization and substrate scope studies. This was accomplished by regressing computationally derived molecular descriptors to the measured enantioselectivities (*36*). Substrates and truncated IDPi catalysts (see supplementary materials) were independently subjected to conformational searches by using molecular mechanics, and the relevant structures were submitted for DFT optimization at the PBE1PBE-D3BJ/def2-SVP level of theory with M06-2X/def2-TZVP single-point energy corrections. Both steric and electronic molecular descriptors were extracted from the lowest energy conformer of the computed ensembles (*37*). The data were split into a training and validation set by using a pseudorandom 50:50 partition followed by forward stepwise multivariate linear regression analysis. A four-parameter model [coefficient of determination (R^2) = 0.85] was found in which (Fig. 4A) two features for both the catalyst and substrate are required for a good correlation (Fig. 4B). Cross-validation techniques, including leave one out (LOO; 0.79) and k-fold ($k = 4$, 0.77), as well as external validation [predicted R^2 ($_{pred}R^2$) = 0.77], were performed to indicate that a statistically robust model was produced. The catalyst is described by the natural bond orbital (NBO) partial charge of the para substituent and the buried volume within a 3.5-Å sphere centered on the sulfonamide sulfur ($VBur_S$), which indicates that a more sterically constricted catalyst pocket positively affects selectivity, consistent with the DFT TS studies. The substrate is described by the ^{13}C nuclear magnetic resonance (NMR) chemical shift and partial equalization of orbital electronegativity (PEOE), which are both primarily electronic descriptors (*38*). The PEOE term can classify general substrate types as well as differentiate the various piperidine protecting groups that were tested during optimization [i.e., *p*-toluenesulfonyl (Ts), Tf, trifluoroacetyl (TFA), and so on] on the basis of the proximity of electronegative atoms (Fig. 4C). Taken together, this model suggests that high enantioselectivity results from a combination of sterically restricted catalysts and dense electronegativity present in the Tf-protected piperidinyl substrates. Catalysis is optimal in the negatively charged pocket of the catalyst for substrates with cation stabilization features, which rationalizes the 93% ee we observed for IDPi (**8B**) with **2f** and supporting the proposed TS.

Extension to strained-ring product motifs

Having established the conceptual groundwork for achieving enantioselectivity in a vinyl cation $C(sp^3)$–H insertion reaction, we sought to test our system with new substrate classes, and we were particularly drawn to the possibility of leveraging the high energy of these intermediates to forge strained rings. Bicyclo[3.2.1]octanes are prevalent carbocyclic scaffolds found in various bioactive natural products (*39*), and we sought to access them in enantioenriched

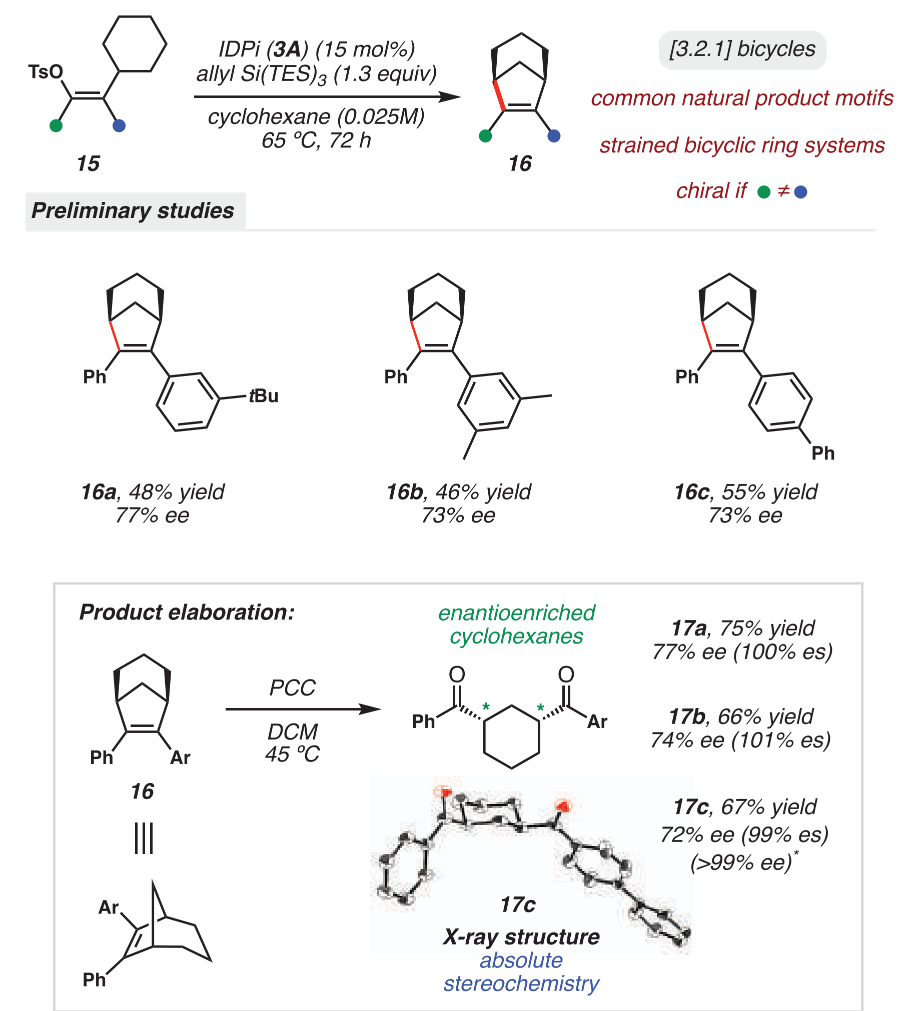


Fig. 5. Catalytic asymmetric synthesis of strained bicycles. All reported yields are isolated yields. *After single recrystallization. Ar, aryl; DCM, dichloromethane; PCC, pyridinium chlorochromate; *t*Bu, *tert*-butyl.

form. To this end, we prepared substrates of the type **15** (Fig. 5), for which insertion into the cyclohexyl group generates a chiral [3.2.1] carbocycle (**16**), and found that IDPi **3A** (Table 1) catalyzes its formation in moderate yields and enantioselectivity (up to 77% ee). Expansion of this chemistry to other substrate classes that generate structurally distinct products suggests the potential for broader applicability of IDPi catalysis in carrying out challenging enantioselective C(sp^3)–H functionalization reactions via vinyl carbocations. Moreover, products from this reaction can be readily converted to enantioenriched 1,3-diketocyclohexanes (**17**) via oxidative cleavage (*40*) with high levels of enantiospecificity. Current catalytic methods to access enantioenriched 1,3-diketocyclohexanes are limited to the desymmetrization of anhydrides (*41*), after which additional steps are required to convert the resulting carboxylic acid to scaffolds such as **17a** to **17c**, which highlights a strategic application of this reaction.

REFERENCES AND NOTES

1. G. A. Olah, *J. Org. Chem.* **66**, 5943–5957 (2001).
2. G. A. Olah, *J. Am. Chem. Soc.* **94**, 808–820 (1972).
3. H. C. Brown, *The Nonclassical Ion Problem* (Plenum Press, 1977).
4. A. Eschenmoser, L. Ruzicka, O. Jeger, D. Arigoni, *Helv. Chim. Acta* **38**, 1890–1904 (1955).
5. D. J. Tantillo, *Nat. Prod. Rep.* **28**, 1035–1053 (2011).
6. R. R. Naredla, D. A. Klumpp, *Chem. Rev.* **113**, 6905–6948 (2013).
7. A. E. Wendlandt, P. Vangal, E. N. Jacobsen, *Nature* **556**, 447–451 (2018).
8. M. Isomura, D. A. Petrone, E. M. Carreira, *J. Am. Chem. Soc.* **141**, 4738–4748 (2019).
9. M. Braun, W. Kotter, *Angew. Chem. Int. Ed.* **43**, 514–517 (2004).
10. K. Brak, E. N. Jacobsen, *Angew. Chem. Int. Ed.* **52**, 534–561 (2013).
11. R. Properzi *et al.*, *Nat. Chem.* **12**, 1174–1179 (2020).
12. P. J. Stang, Z. Rappoport, *Dicoordinated Carbocations* (Wiley, 1997).
13. B. Shao, A. L. Bagdasarian, S. Popov, H. M. Nelson, *Science* **355**, 1403–1407 (2017).
14. L. Mascarelli, *Gazz. Chim. Ital.* **66**, 843–850 (1936).
15. S. Popov *et al.*, *Science* **361**, 381–387 (2018).
16. B. Wigman *et al.*, *J. Am. Chem. Soc.* **141**, 9140–9144 (2019).
17. S. E. Cleary, M. J. Hensinger, M. Brewer, *Chem. Sci.* **8**, 6810–6814 (2017).
18. U. Biermann, R. Koch, J. O. Metzger, *Angew. Chem. Int. Ed.* **45**, 3076–3079 (2006).
19. F. Zhang *et al.*, *J. Am. Chem. Soc.* **136**, 8851–8854 (2014).
20. S. E. Cleary *et al.*, *J. Am. Chem. Soc.* **141**, 3558–3565 (2019).
21. I. M. Riddlestone, A. Kraft, J. Schaefer, I. Krossing, *Angew. Chem. Int. Ed.* **57**, 13982–14024 (2018).
22. C. K. De, R. Mitra, B. List, *Synlett* **28**, 2435–2438 (2017).
23. P. Pommerening, M. Oestreich, *Eur. J. Org. Chem.* **2019**, 7240–7246 (2019).
24. K. U. Wendt, A. Lenhart, G. E. Schulz, *J. Mol. Biol.* **286**, 175–187 (1999).
25. B. Mitschke, M. Turberg, B. List, *Chem* **6**, 2515–2532 (2020).
26. C. M. Hong, R. G. Bergman, K. N. Raymond, F. D. Toste, *Acc. Chem. Res.* **51**, 2447–2455 (2018).
27. P. S. J. Kaib, L. Schreyer, S. Lee, R. Properzi, B. List, *Angew. Chem. Int. Ed.* **55**, 13200–13203 (2016).
28. L. Schreyer, R. Properzi, B. List, *Angew. Chem. Int. Ed.* **58**, 12761–12777 (2019).
29. E. Fattorusso, O. Taglialatela-Scafati, *Modern Alkaloids: Structure, Isolation, Synthesis and Biology* (Wiley, 2007).
30. J. B. Lambert, S. Zhang, S. M. Ciro, *Organometallics* **13**, 2430–2443 (1994).
31. K. Miyamoto, M. M. Hoque, S. Ogasa, *J. Org. Chem.* **77**, 8317–8320 (2012).
32. A. J. Neel, A. Milo, M. S. Sigman, F. D. Toste, *J. Am. Chem. Soc.* **138**, 3863–3875 (2016).
33. H. Zhou *et al.*, *J. Am. Chem. Soc.* **142**, 13695–13700 (2020).
34. C. C. Wagen, E. E. Kwan, *cctk* (2020); www.github.com/ekwan/cctk.
35. J. S. Dickschat, *Nat. Prod. Rep.* **33**, 87–110 (2016).
36. C. B. Santiago, J.-Y. Guo, M. S. Sigman, *Chem. Sci.* **9**, 2398–2412 (2018).
37. J.-Y. Guo, Y. Minko, C. B. Santiago, M. S. Sigman, *ACS Catal.* **7**, 4144–4151 (2017).
38. J. Gasteiger, M. Marsili, *Tetrahedron* **36**, 3219–3228 (1980).
39. M. Presset, Y. Coquerel, J. Rodriguez, *Chem. Rev.* **113**, 525–595 (2013).
40. M. Miao *et al.*, *Org. Lett.* **21**, 5957–5961 (2019).
41. M. J. Cook, T. Rovis, *J. Am. Chem. Soc.* **129**, 9302–9303 (2007).

ACKNOWLEDGMENTS

We thank S. Virgil, B. Stoltz, and S. Reisman for instrumentation; M. Shahgholi for mass spectrometry support; D. Vander Velde for NMR expertise; and M. Takase and S. Khan for x-ray crystallography analysis. Work by the Houk group used computational and storage services associated with the Hoffman2 Shared Cluster provided by UCLA Institute for Digital Research and Education's Research Technology Group. The Sigman Group's computational work was supported by the Center for High Performance Computing at the University of Utah. **Funding:** Financial support for this work was provided by the National Institutes of Health National Institute of General Medical Sciences (R35 GM12893 to H.M.N. and R35 GM136271 to M.S.S.) and the National Science Foundation (CHE-1764328 to K.N.H.; DGE-1650604 to B.W. and S.P.). S.K.N. thanks the US Public Health Service (USPHS) National Research Service Award (T32GM008496). We thank the Beckman Institute (support of the Caltech Center for Catalysis and Chemical Synthesis and the x-ray crystallography facility), the Dow Next-Generation Educator Fund (grant to Caltech), and the UCLA Molecular Instrumentation Center for NMR instrumentation and x-ray crystallography. **Author contributions:** H.M.N., S.K.N., C.G.W., B.W., and S.P. designed and conducted experiments. Computational studies were conducted by J.J.W. and directed by K.N.H. Statistical modeling of data was performed by B.C.H. and J.W., and M.S.S. supervised. H.M.N., M.S.S., S.K.N., C.G.W., B.W., B.C.H., J.J.W., and J.W. prepared the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All crystallographic data are available free of charge from the Cambridge Crystallographic Data Centre under CCDC 2201595, CCDC 2201596, CCDC 2201597, CCDC 2201598, CCDC 2201599, and CCDC 2201600. All other data are provided in the supplementary materials. **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.ade5320
Materials and Methods
Figs. S1 to S24
Tables S1 to S8
Spectral Data
References (*42*–*102*)
Data S1

Submitted 22 August 2022; accepted 8 November 2022
10.1126/science.ade5320

COMPUTER SCIENCE

Competition-level code generation with AlphaCode

Yujia Li*†, David Choi*†, Junyoung Chung†, Nate Kushman†, Julian Schrittwieser†, Rémi Leblond†, Tom Eccles†, James Keeling†, Felix Gimeno†, Agustin Dal Lago†, Thomas Hubert†, Peter Choy†, Cyprien de Masson d'Autume†, Igor Babuschkin, Xinyun Chen, Po-Sen Huang, Johannes Welbl, Sven Gowal, Alexey Cherepanov, James Molloy, Daniel J. Mankowitz, Esme Sutherland Robson, Pushmeet Kohli, Nando de Freitas, Koray Kavukcuoglu, Oriol Vinyals*

Programming is a powerful and ubiquitous problem-solving tool. Systems that can assist programmers or even generate programs themselves could make programming more productive and accessible. Recent transformer-based neural network models show impressive code generation abilities yet still perform poorly on more complex tasks requiring problem-solving skills, such as competitive programming problems. Here, we introduce AlphaCode, a system for code generation that achieved an average ranking in the top 54.3% in simulated evaluations on recent programming competitions on the Codeforces platform. AlphaCode solves problems by generating millions of diverse programs using specially trained transformer-based networks and then filtering and clustering those programs to a maximum of just 10 submissions. This result marks the first time an artificial intelligence system has performed competitively in programming competitions.

Automatically creating programs given a high-level description of what they should do is a long-standing task in computer science (*1*, *2*). Creating an artificial intelligence (AI) system that can solve unforeseen problems by generating code from problem descriptions is a challenge that both affords a greater understanding of problem solving and reasoning (*3*) and leads to important applications, such as improving programmer productivity (*4*) and education (*5*).

Generating code that solves a specified task requires searching in the enormous space of all possible character sequences, only a tiny portion of which corresponds to valid and correct programs. Furthermore, single character edits can completely change program behavior or even cause crashes, and each task has many valid solutions that may be drastically different. These challenges make learning to generate correct programs difficult. Most prior work has been limited to either restricted domain-specific programming languages (*6*) or short code snippets (*7*, *8*). Perhaps the best-known examples of program synthesis are Flash Fill (*6*), which synthesizes programs from input and output examples to automatically fill in data in Microsoft Excel, and code completion tools common in integrated development environments, which boost programmer productivity (*9*, *10*).

Contrasting with the conceptually more complex systems used in most of program synthesis history, recent large-scale transformer-based (*11*) language models [which achieve impressive performance generating text (*12*)] can, with minimal modification, solve simple programming problems in Python (*13*, *14*). A stripped-down version of our model performs similarly to these prior works (table S13). However, those problems consist mostly of simple task descriptions with short solutions, and solving them often amounts to translating a sequence of steps (e.g., adding together all even numbers in a list) directly into code. In contrast, generating entire programs often relies on understanding the task (e.g., win a board game), reasoning out the appropriate algorithm to solve it, and then writing the code to implement that algorithm.

Solving competitive programming problems (Fig. 1A) represents a big step forward. It requires understanding complex natural language descriptions, reasoning about previously unseen problems instead of simply memorizing code snippets, mastering a wide range of algorithms and data structures, and precisely implementing submissions that can span hundreds of lines. To evaluate these submissions (Fig. 1B), they are executed on an exhaustive suite of hidden tests and checked for execution speed and correctness on edge cases. Feedback is minimal; the submission is correct only if it has the correct output on all hidden tests, otherwise it is incorrect. Hidden tests are not visible to the submitter, who must instead write their own tests or rely on the trivial example tests for debugging. Because competitors are allowed to draw on solutions and algorithms from previous contests, challenging new problems are created for each competition. Competitive programming is very popular; events such as the International Collegiate Programming Competition (*15*) and the International Olympiad in Informatics (*16*) date back to the 1970s and are some of the most prestigious competitions in computer science, drawing hundreds of thousands of participants from around the world. Using problems that humans find challenging from battle-tested competitions provides a robust and meaningful benchmark for many aspects of intelligence.

Early work using program synthesis for competitive programming has shown that large transformer models can achieve low single-digit solve rates (*13*, *17*). In contrast, we created a code generation system named AlphaCode that manages to solve 29.6% of test set held-out competitive programming problems in a dataset we released named CodeContests, using at most 10 submissions per problem (comparable to humans). A key driver of AlphaCode's performance came from scaling the number of model samples to orders of magnitude more than previous work; the overall solve rate scaled log-linearly with the number of samples generated, even when only 10 of them were submitted, a sample scaling law similar to those found for training compute and model size (*12*).

When evaluated on simulated programming competitions hosted on the popular Codeforces platform (*18*), AlphaCode achieved an average ranking within the top 54.3% of human participants (a small, selected subset of all programmers). To the best of our knowledge, a computer system has never before achieved such a competitive level in programming competitions.

Learning system

Our system (Fig. 2) was designed to address the main challenges of competitive programming: (i) searching in the huge space of programs, (ii) access to only ~13,000 example tasks

Table 1. AlphaCode's results on Codeforces competitions. For each contest, we show the estimated percent ranking (lower is better) using simulated time and incorrect submission penalties, as well as the best and worst possible rankings using minimum and maximum time penalties, averaged over three evaluations. Percents are how many users performed better than AlphaCode. AlphaCode achieved an overall ranking in the top 54.3% averaged across the 10 contests.

Contest ID	1591	1608	1613	1615	1617	1618	1619	1620	1622	1623	Average
Maximum	43.5%	43.6%	59.8%	60.5%	65.1%	32.2%	47.1%	54.0%	57.5%	20.6%	48.4%
Estimated	44.3%	46.3%	66.1%	62.4%	73.9%	52.2%	47.3%	63.3%	66.2%	20.9%	54.3%
Minimum	74.5%	95.7%	75.0%	90.4%	82.3%	53.5%	88.1%	75.1%	81.6%	55.3%	77.2%

DeepMind, London, UK.
*Corresponding author. Email: yujiali@deepmind.com (Y.L.); davidhchoi@deepmind.com (D.C.); vinyals@deepmind.com (O.V.)
†These authors contributed equally to this work.

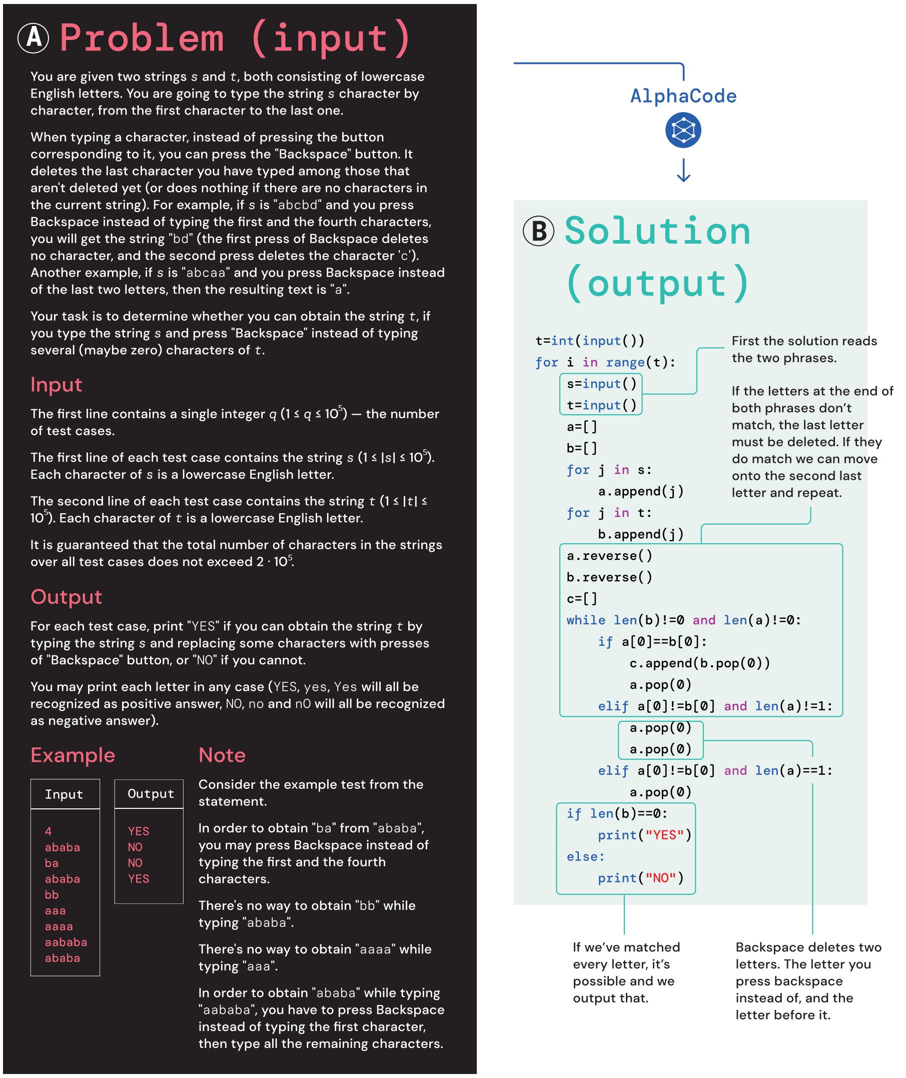


Fig. 1. Competitive programming problem statement and solution. (**A**) Problem statement of 1553D ("Backspace"; https://codeforces.com/problemset/problem/1553/D), a Codeforces problem (*18*). This is a medium-difficulty problem, with a rating of 1500, and its statement includes a public example test case. The entire statement is given to AlphaCode. (**B**) Solution to the problem statement, generated by AlphaCode. Examples of the exact formatting of problem descriptions, solutions, and what the model perceives can be found at https://alphacode.deepmind.com (*32*).

for training, and (iii) a restricted number of submissions per problem. Notably, scaling the amount of model samples generated greatly improved performance (Fig. 3), and therefore many aspects of our system were designed to draw samples as quickly as possible, to ensure sample diversity, and to select the best samples for submission. See supplementary materials (SM) text section C for full details of the method.

Creating solutions to problems can be seen as a sequence-to-sequence (*19*) prediction task: Given a problem description X in natural language (e.g., Fig. 1A), produce a corresponding solution Y in a programming language (e.g., Fig. 1B). This task naturally motivated an encoder-decoder transformer architecture (*11*) for AlphaCode, which models $p(Y|X)$. The architecture took as input to the encoder the problem description and metadata X as a flat sequence of tokenized (*20*) characters and sampled Y autoregressively from the decoder one token at a time until an end-of-code token was produced. The code was then ready to be compiled and run against tests.

We pretrained our models on a 715-GB snapshot of human code from GitHub (*21*), with cross-entropy next-token prediction loss. During pretraining, we randomly split code files into two parts, using the first part as input to the encoder and training the decoder to produce the second part. This pretraining learned a strong prior for human coding, enabling subsequent task-specific fine-tuning with a much smaller dataset.

We fine-tuned and evaluated models on a 2.6-GB dataset of competitive programming problems that we created and released under the name CodeContests (*22*). CodeContests includes problems, correct and incorrect human submissions, and test cases. The training set contains 13,328 problems, with an average of 922.4 submissions per problem. The validation and test sets contain 117 and 165 problems, respectively. Full sets of hidden test cases used in competitions were not available, so we included extra generated tests in CodeContests (created by mutating existing tests) to decrease the false positive rate of our model's submissions from 60% (comparable to other datasets) to 4%. We still observed that 42% of solutions were correct but algorithmically inefficient, running out of time or memory on larger test inputs, which motivated the crucial step of evaluating model submissions on the official competitive programming website for the main evaluation. The validation set of CodeContests was used as an easy-to-measure, low-variance proxy of submitting to a competition, and the test set was used only for final evaluations after all training and tuning had been completed. All data used to train the models (in pretraining and fine-tuning) appeared online before all problems in the validation and test sets, ensuring that only information that humans could have seen before a competition would have been available to the models.

During fine-tuning, we encoded the natural language problem statement as a program comment, to make it appear more similar to files seen during pretraining (which can include extended natural language comments), and used the same next-token prediction loss. We trained a variety of models ranging from 300 million to 41 billion parameters.

Our models included the following enhancements on top of a standard transformer encoder-decoder architecture. Each improvement improved the solve rate, as we show in the evaluation section:

1) Multi-query attention: Multi-query attention (*23*) used a full set of query heads but shared key and value heads per attention block to substantially reduce memory usage and cache updates (which were bottlenecks during sampling). In combination with an encoder-decoder model instead of a decoder-only one, the change increased the sampling speed of AlphaCode more than 10-fold.

2) Masked language modeling (MLM): We included an MLM loss (*24*) on the encoder, which empirically improved model solve rates.

3) Tempering: Tempering (*25*) artificially made the training distribution sharper, producing a smoother sampling distribution (a regularization effect to prevent overfitting).

4) Value conditioning and prediction: We used value conditioning and prediction to discriminate between correct and incorrect

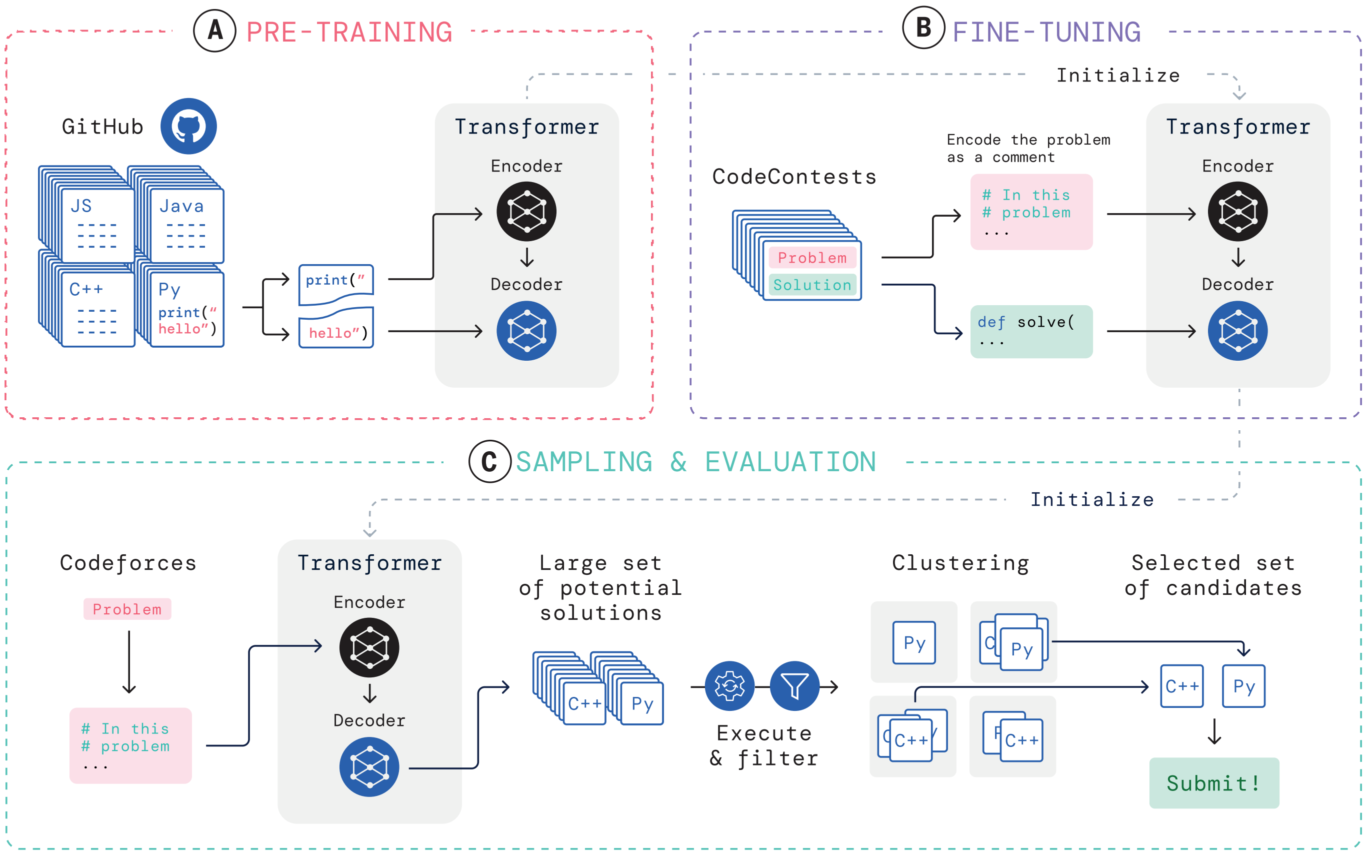


Fig. 2. Overview of AlphaCode. (**A**) In pretraining, files from GitHub are randomly split into two parts. The first part goes to the encoder as input, and the decoder is trained to produce the second part. (**B**) In fine-tuning, problem descriptions (formatted as comments) are given to the encoder, and the decoder is trained to generate the solutions. (**C**) For evaluation, AlphaCode generates many samples for each problem description, then it executes them to filter out bad samples and cluster the remaining ones before finally submitting a small set of candidates.

problem submissions contained in CodeContests, providing an additional training signal and allowing the use of incorrect submission data that might otherwise mislead the model.

5) Generation by off-policy learning from demonstrations (GOLD): The CodeContests training set contains hundreds of solutions for each problem. The standard cross-entropy next-token prediction loss would put equal weight on all solutions. A successful model, however, only needs to generate a single correct solution for each problem. To resolve this discrepancy, we adopted a variation of GOLD (*26*), an offline reinforcement learning (RL) algorithm that focuses training on the most likely solutions for each problem instead of all possible solutions.

At sampling time, the diversity of samples was important, so that the millions of samples for each problem could effectively explore the space of possible solutions. As in another publication (*27*), we ensured sample diversity by using a high temperature and conditioning samples on random metadata: problem difficulty ratings, problem tags (which indicate which techniques a solution might use), and solution programming language.

To select the 10 best samples for submission, we applied filtering and clustering to obtain a small number of candidate submissions on the basis of their program behavior. Filtering executed samples using example tests included in the problem statement and removed samples that failed those tests. This filtering removed ~99% of model samples. The possibly tens of thousands of candidate samples that remained were then clustered by executing them on inputs generated by a separate transformer model trained to do test input generation and by grouping together programs that produced the same outputs on the generated inputs. We then picked one sample from each of the 10 largest clusters for submission, approximately the most likely program behaviors from our model. Intuitively, correct programs would behave the same and form large clusters, but incorrect programs could fail in many different ways.

Evaluation

To assess the performance of AlphaCode, we evaluated it against programming competitions from the Codeforces platform (*18*). Compared with reporting the solve rate on a dataset, this evaluation avoids dataset assumptions and weaknesses that could skew results and allows us to benchmark against the best performers on this task–human competitors.

We ensembled our 41 billion (41B) and 9 billion (9B) parameter models by pooling their samples and then evaluated the ensemble on all Codeforces competitions from 1 December 2021 to 28 December 2021 with >5000 participants per contest, a total of 10 competitions that we believe are a representative sample of Codeforces contests. For each contest, we simulated running AlphaCode live, generating samples for each problem, filtering with example tests (*28*), and clustering to get candidate submissions. We submitted these candidates to the Codeforces platform and computed AlphaCode's placement in each contest (Table 1). After the first run, we repeated this procedure two more times to measure variance.

Overall, our system achieved an average ranking in the top 54.3% when limited to 10 submissions per problem, although 66% of solved problems were solved with the first submission. This performance in competitions approximately corresponds to a novice programmer with a few months to a year of training (see SM text section G). To the best of our knowledge,

Table 2. Buildup ablation for model enhancements. Effect of each additional model enhancement building up from "No enhancements"—a plain fine-tuned 1B encoder-decoder model trained with the standard next-token prediction loss. Numbers in parentheses represent 95% confidence intervals. For each setting, we fine-tuned and sampled from at least three different models from the same pretrained checkpoint.

	Solve rate	
Fine-tuning setting	**10@1K**	**10@1M**
No enhancements	6.7% (6.5–6.8)	19.6% (18.2–20.4)
+ MLM	6.6% (6.2–7.0)	20.7% (19.1–21.3)
+ Tempering	7.7% (7.2–8.5)	21.9% (20.7–22.6)
+ Random tags and ratings	6.8% (6.4–7.0)	22.4% (21.3–23.0)
+ Value	10.6% (9.8–11.1)	23.2% (21.7–23.9)
+ GOLD	12.4% (12.0–13.0)	24.2% (23.1–24.4)
+ Clustering	12.2% (10.8–13.4)	28.4% (27.5–29.3)

this is the first time that a computer system has been competitive with human participants in programming competitions. Training and evaluating our largest 41B model on Codeforces required a total of 2149 petaflop/s-days and 175 megawatt-hours [~16 times the average American household's yearly energy consumption (*29*)]. The large resource usage required for this experiment both has environmental impacts and makes replication challenging for most research entities.

To assess our modeling decisions, we evaluated our models on the validation and test sets of CodeContests. The main CodeContests metric was 10@*k*: the percentage of problems solved when we take *k* samples from the model for each problem but can only submit 10 of them for evaluation on the hidden tests. If any of the 10 selected samples solved a problem, the problem is solved. This limit of 10 submissions mimics the upper bound of the number of submissions a human would typically use, and tracking *k* is important for comparisons, as performance improves as *k* increases.

With up to 100,000 samples per problem (10@100K), the AlphaCode 41B model solved 29.6% of problems in the CodeContests test set. Prior work on comparable datasets achieved low single-digit solve rates (*13*, *17*), and direct comparisons can be found in SM text section E.3.

Figure 3 shows how the model performance scaled on the 10@*k* metrics with more samples, that is, as we increased *k*, or with more compute. These scaling curves highlight a few notable facts about this problem domain and our models:

1) Solve rates scale log-linearly with more samples. As shown in Fig. 3A, the 10@*k* solve rates scaled approximately log-linearly with *k*, bending down slightly at high sample budgets. The fact that sampling much more than 10 still improved the 10@*k* solve rate shows how important it was to sufficiently explore the search space before committing to the final 10 submissions per problem. However, improving the solve rate required exponentially increasing amounts of samples, and the costs quickly became prohibitive.

2) Better models have higher slopes in the scaling curve. Figure 3A also shows that larger models tended to generate higher-quality samples, reflected as better solve rates with the same number of samples and higher slopes in this log-linear scaling curve. Because of log-linear scaling, a better model could reach the same solve rates with exponentially fewer samples than worse models. Improving model quality is an effective way to counter the exponential explosion of sample budget required for a higher solve rate.

3) Solve rates scale log-linearly with more compute. As shown in Fig. 3B, the solve rate also scaled approximately log-linearly with more training compute. Each vertical slice on the curves corresponds to one model size. Figure 3C shows how the solve rate scaled with sampling compute: Larger models required more compute per sample but eventually outperformed smaller models, as the better quality of samples from the larger models became the dominant factor for performance. Both ways of leveraging more compute demonstrate log-linear scaling.

4) Sample selection is critical to solve rate scaling. Figure 3D shows how 10@*k* solve rates scaled when we used different sample selection methods. With clustering and filtering, as was used in AlphaCode, taking more samples improved the solve rate. However, the combination was far from the theoretical upper bound of performance with perfect sample selection. Without filtering (bottom curve), 10 samples were randomly selected for submission, and taking more samples did not improve the solve rate.

Table 2 shows a buildup ablation of AlphaCode's enhancements, which significantly improved the solve rate on models at the 1B scale. We started from the base setting with no enhancements (beyond the multi-query attention change). We added one new setting at a time, with the final setting that corresponds to AlphaCode reported at the bottom of the table. Combining the six enhancements together increased the 10@1M solve rate from 19.6% to 28.4%, although the improvement depended on the number of samples.

Model analysis

To explore the possibility of solving problems by exploiting weaknesses in the problem structure, we analyzed the capabilities and limitations of our models. A common concern for large language models trained on large amounts of data is that they solve downstream problems by simply memorizing the training set (*30*, *31*). For competitive programming to be a good test of problem-solving ability, models need to come up with novel solutions to solve new problems. We found no evidence that our model copied core logic from the training data. Copying solutions from the training set was not sufficient to solve any problems in the unseen validation set, and the longest substrings that both human solutions and AlphaCode solutions shared with the training data were of similar lengths (fig. S10). Further manual investigation showed that core logic of model solutions necessary to solve problems could not be found in the training data.

We also analyzed the properties of model samples. Like humans, models correctly solved easier problems more often than harder problems; the 41B model solved 62.4% of validation problems rated between 800 and 1100 and 7.8% of problems rated between 2400 and 2700 (higher-rated problems are more difficult). Models had the highest solve rates for problems that deal with bitmasks, sorting, greedy algorithms, or math, but the lowest solve rates for dynamic programming, constructive algorithm, and graph problems. The former categories tend to afford shorter solutions and the latter categories afford longer ones, indicating that AlphaCode had trouble producing longer solutions.

AlphaCode was sensitive to changes in the problem descriptions, indicating that it did not exploit obvious weaknesses in the task structure. We modified a problem statement that asked for a program that can find the maximum product of two consecutive array elements. When AlphaCode generated samples for the opposite problem (minimum instead of maximum product) and a related problem (any elements instead of consecutive ones), the percentage correctness of those samples on the original problem decreased from 17.1% to 0.1% and 3.2% for the opposite and related rewordings, respectively. AlphaCode attended to important changes in problem

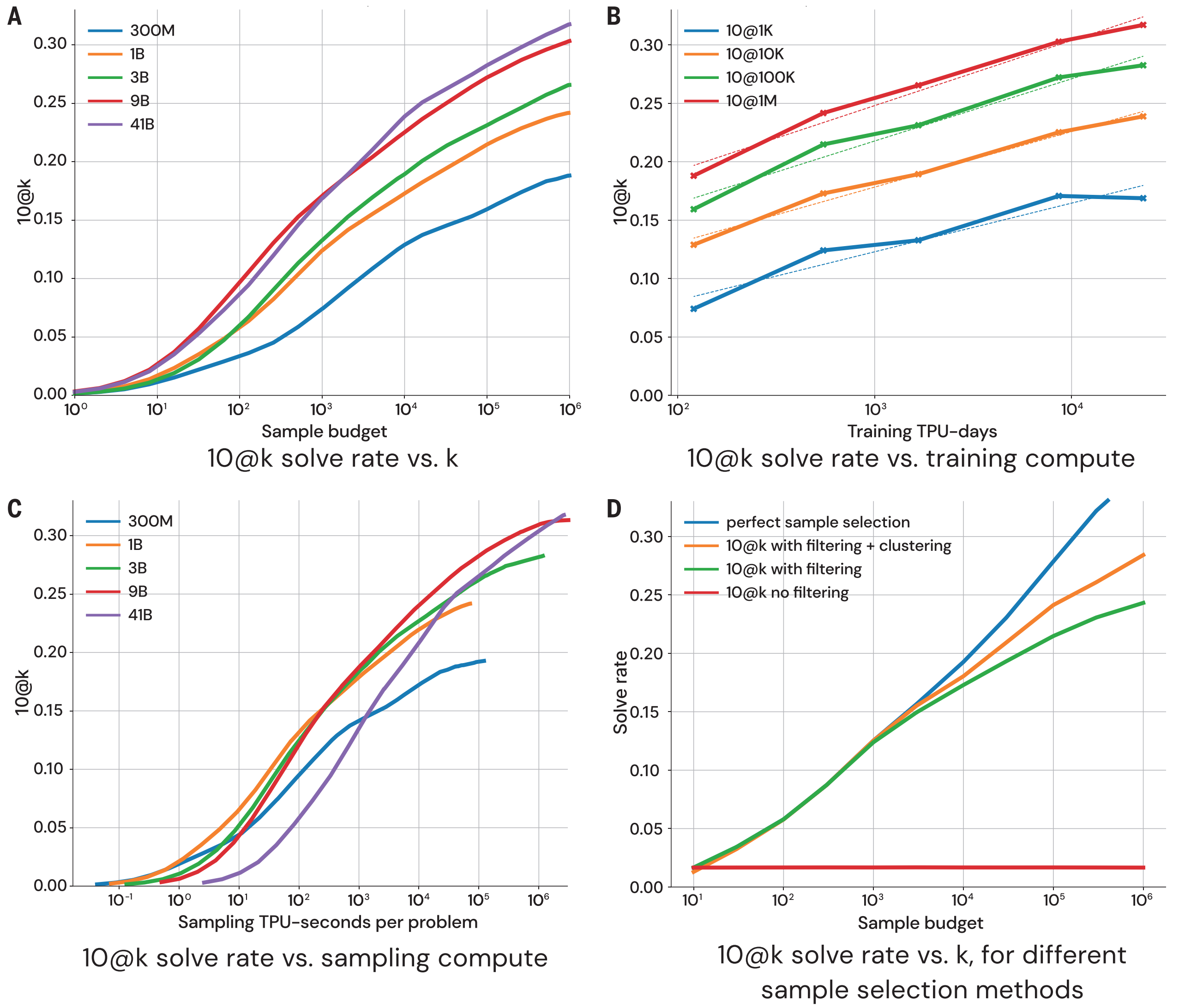


Fig. 3. Solve rate scaling. (**A**) The solve rate scaled approximately log-linearly with the number of samples, tapering off slightly in the 10@*k* setting. Larger models with more parameters had higher scaling slopes in this log-linear plot. (**B**) The solve rate scaled approximately log-linearly with the training compute when we choose near-optimal model sizes for each compute allocation. (**C**) As we increased the amount of compute we used for sampling, the optimal model size increased. (**D**) Improved sample selection methods were necessary to see solve rate scaling. In order of quality: random selection ("10@*k* no filtering"), filtering using example tests ("10@*k* with filtering"), clustering after filtering ("10@*k* with filtering + clustering"), and perfect sample selection. TPU, tensor processing unit.

descriptions that fundamentally changed the problem, instead of uniformly trying every related algorithm by abusing the large sample budget. Conversely, alterations that fundamentally left the problem unchanged affected solve rates less, indicating that AlphaCode could mostly ignore changes that humans could also ignore. For example, introducing 30 adjacent character transposition typos decreased the 10@1024 solve rate from 13.5% to 11.3%. More analysis can be found in SM text section F.

Conclusion

Here, we present AlphaCode, a system applied to code generation for competitive programming that can generate novel solutions to unseen programming problems. When evaluated on Codeforces, AlphaCode performed roughly at the level of the median competitor. We found that massively scaling up sampling and then filtering and clustering samples to a small set, together with efficient transformer architectures to support large-scale sampling, were essential to achieving good performance. Our clean dataset and robust evaluation procedure [released as CodeContests (*22*)] also contributed substantially to guiding our research progress. We show through detailed analysis that there is no evidence that AlphaCode copied important parts of previous solutions or exploited weaknesses in the problem structure. The analysis indicates that our model was indeed able to solve problems it has never seen before, even though those problems require considerable reasoning.

Combined with careful research and engineering, the relatively simple architecture of transformers shows exceptional potential in performing reasoning necessary to solve complex problems with code. This line of work has exciting applications that can improve the productivity of programmers and make programming accessible to a new generation of developers. Our work in code generation still leaves much room for improvement, and we hope that further work will advance the fields of both programming and reasoning in AI.

REFERENCES AND NOTES

1. Z. Manna, R. J. Waldinger, *Commun. ACM* **14**, 151–165 (1971).
2. A. Church, *J. Symb. Log.* **28**, 289–290 (1963).
3. C. C. Green, in *Readings in Artificial Intelligence*, B. L. Webber, N. J. Nilsson, Eds. (Elsevier, 1981), pp. 202–222.

4. N. D. Matsakis, F. S. Klock II, *ACM SIGAda Ada Lett.* **34**, 103–104 (2014).
5. M. Resnick *et al.*, *Commun. ACM* **52**, 60–67 (2009).
6. S. Gulwani, *SIGPLAN Not.* **46**, 317–330 (2011).
7. V. Raychev, M. Vechev, E. Yahav, *SIGPLAN Not.* **49**, 419–428 (2014).
8. M. Bruch, M. Monperrus, M. Mezini, "Learning from examples to improve code completion systems," *Proceedings of the 7th Joint Meeting of the European Software Engineering Conference and the ACM SIGSOFT Symposium on the Foundations of Software Engineering (ESEC/FSE '09)*, Amsterdam, Netherlands, 24 to 28 August 2009 (Association for Computing Machinery, 2009), pp. 213–222.
9. A. Hindle, E. T. Barr, Z. Su, M. Gabel, P. Devanbu, "On the naturalness of software," *Proceedings of the 34th International Conference on Software Engineering (ICSE '12)*, Zurich, Switzerland, 2 to 9 June 2012 (IEEE Press, 2012), pp. 837–847.
10. A. Svyatkovskiy, S. K. Deng, S. Fu, N. Sundaresan, "IntelliCode compose: code generation using transformer," *Proceedings of the 28th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering(ESEC/FSE '20)*, 8 to 13 November 2020 (Association for Computing Machinery, 2020), pp. 1433–1443.
11. A. Vaswani *et al.*, *Adv. Neural Inf. Process. Syst.* **30**, 5998–6008 (2017).
12. T. B. Brown *et al.*, arXiv:2005.14165 [cs.CL] (2020).
13. M. Chen *et al.*, arXiv:2107.03374 [cs.LG] (2021).
14. J. Austin *et al.*, arXiv:2108.07732 [cs.PL] (2021).
15. International Collegiate Programming Contest (ICPC), https://icpc.global/.
16. International Olympiad in Informatics (IOI), https://ioinformatics.org/.
17. D. Hendrycks *et al.*, arXiv:2105.09938 [cs.SE] (2021).
18. Codeforces, https://codeforces.com/.
19. I. Sutskever, O. Vinyals, Q. V. Le, *Adv. Neural Inf. Process. Syst.* **27**, 3104–3112 (2014).
20. T. Kudo, J. Richardson, arXiv:1808.06226 [cs.CL] (2018).
21. The public GitHub dataset is hosted on Google BigQuery at https://console.cloud.google.com/marketplace/product/github/github-repos.
22. The released CodeContests dataset can be found at https://github.com/deepmind/code_contests and at https://doi.org/10.5281/zenodo.6975437.
23. N. Shazeer, arXiv:1911.02150 [cs.NE] (2019).
24. J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* (Association for Computational Linguistics, 2019), pp. 4171–4186.
25. R. Dabre, A. Fujita, arXiv:2009.09372 [cs.CL] (2020).
26. R. Y. Pang, H. He, arXiv:2009.07839 [cs.CL] (2020).
27. O. Vinyals *et al.*, *Nature* **575**, 350–354 (2019).
28. For problems permitting multiple correct outputs, we changed the example test outputs to be the most canonical, which gives our approach a slight advantage in the evaluation. See SM text section D for more details.
29. US Energy Information Administration, Electric Sales, Revenue, and Average Price: Summary Table T5.a: 2021 Residential Average Monthly Bill by Census Division, and State; https://www.eia.gov/electricity/sales_revenue_price/.
30. A. Ziegler, "GitHub Copilot research recitation: GitHub Copilot: Parrot or Crow? A first look at rote learning in GitHub Copilot suggestions," The GitHub Blog, 30 June 2021; https://docs.github.com/en/github/copilot/research-recitation.
31. N. Carlini *et al.*, "Extracting training data from large language models," *30th USENIX Security Symposium (USENIX Security 21)*, 11 to 13 August 2021, pp. 2633–2650.
32. AlphaCode examples, explanations, and visualizations can be found at https://alphacode.deepmind.com and at https://doi.org/10.5281/zenodo.6975437.
33. Y. Li *et al.*, AlphaCode data materials, version 1.0.0, Zenodo (2022); https://doi.org/10.5281/zenodo.6975437.

ACKNOWLEDGMENTS

We thank T. Cai, J. Rae, S. Borgeaud, M. Glaese, R. Ring, L. Sifre, J. Hoffman, J. Aslanides, J.-B. Lespiau, A. Mensch, E. Elsen, G. van den Driessche, and G. Irving for developing tools that we use to train large language models and for lending their expertise in model training; K. Anderson, C. Pope, and R. Foley for project management in early stages; Y. Whye Teh, C. Dyer, D. Silver, A. Barekatain, A. Zhernov, M. Overlan, and P. Veličković for research advice and assistance; K. Simonyan, C. Dyer, and D. Yogatama for reviewing the paper; L. Bennett, K. Ayoub, and J. Stanway for logistically making the project possible; S. Dathathri for analyzing our model; E. Caballero for granting permission to use Description2Code data; R. Ke for helping connect us with E. Caballero; P. Heiber for helping connect us with Codeforces; P. Mitrichev for helping connect us with Codeforces, and lending competitive programming expertise as the paper was being written; M. Mirzayanov for allowing us to evaluate on Codeforces and for lending competitive programming expertise when writing the paper; and everyone at DeepMind for their insight and support. **Funding:** All research in this paper was funded by DeepMind and Alphabet. There was no external funding. **Author contributions:** Y.L. and D.C. led the project; C.d.M.d'A., D.J.M., D.C., F.G., J.K., J.S., J.C., N.K., P.C., R.L., T.H., T.E., X.C., and Y.L. designed and implemented the learning system and algorithm; J.C. and N.K. ran the final experiments; A.D.L., D.C., F.G., J.K., J.M., J.S., J.C., N.K., P.C., R.L., T.H., T.E., and Y.L. worked on infrastructure for running and evaluating experiments; A.C., C.d.M.d'A., J.W., N.K., P.C., P.-S.H., R.L., S.G., and T.E. worked on model analysis; A.D.L., D.C., F.G., J.S., J.C., N.K., R.L., T.E., and Y.L. worked on datasets; D.C., I.B., J.C., N.K., and Y.L. worked on initial prototypes and infrastructure; D.C., J.K., J.S., J.C., N.d.F., N.K., O.V., P.C., R.L., T.E., and Y.L. wrote the paper; D.C., E.S.R., O.V., and Y.L. worked on project management; and N.d.F., O.V., K.K., and P.K. advised the project. **Competing interests:** This work was done in the course of employment at DeepMind, with no other competing financial interests. DeepMind has filed the following patent application related to this work: US63/306,043. **Data and materials availability:** The datasets used in the experiments as well as evaluation code have been made available for download on GitHub at https://github.com/deepmind/code_contests. Visualizations of the AlphaCode model, selected samples, and explanations can be found at https://alphacode.deepmind.com. Both of the above, as well as results of manual marking, can also be found in Zenodo (*33*). All other data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials.

License information: Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abq1158
Supplementary Text
Figs. S1 to S29
Tables S1 to S23
References (*34*–*88*)

Submitted 29 April 2022; accepted 4 October 2022
10.1126/science.abq1158

REPORTS

CHEMICAL GENETICS

IFITM proteins assist cellular uptake of diverse linked chemotypes

Kevin Lou[1,2], Douglas R. Wassarman[1,2], Tangpo Yang[1], YiTing Paung[3], Ziyang Zhang[1,2,4], Thomas A. O'Loughlin[5,6], Megan K. Moore[1,2], Regina K. Egan[7], Patricia Greninger[7], Cyril H. Benes[7,8], Markus A. Seeliger[3], Jack Taunton[1], Luke A. Gilbert[5,6,9,10]*, Kevan M. Shokat[1,2,4]*

The search for cell-permeable drugs has conventionally focused on low-molecular weight (MW), nonpolar, rigid chemical structures. However, emerging therapeutic strategies break traditional drug design rules by employing flexibly linked chemical entities composed of more than one ligand. Using complementary genome-scale chemical-genetic approaches we identified an endogenous chemical uptake pathway involving interferon-induced transmembrane proteins (IFITMs) that modulates the cell permeability of a prototypical biopic inhibitor of MTOR (RapaLink-1, MW: 1784 g/mol). We devised additional linked inhibitors targeting BCR-ABL1 (DasatiLink-1, MW: 1518 g/mol) and EIF4A1 (BisRoc-1, MW: 1466 g/mol), uptake of which was facilitated by IFITMs. We also found that IFITMs moderately assisted some proteolysis-targeting chimeras and examined the physicochemical requirements for involvement of this uptake pathway.

Any therapeutic molecule that binds to an intracellular target must first cross the cell membrane. Retrospective analyses of compound libraries and their biological activities have yielded empirical guidelines (e.g., Lipinski's rule of five) that enrich for lead-like scaffolds with high passive permeability and largely define modern drug-like chemical space (*1*–*3*). Although these principles have been useful for streamlining the search for new therapeutics, many important intracellular drug targets are currently refractory to inhibition by these compact, hydrophobic, and rigid molecules. One emerging design framework that seeks to address these challenges involves increasing pharmacological complexity by linking multiple ligands into a single chemical entity (a linked chemotype). Linked chemotypes can have enhanced potency, greater selectivity, and the capacity to induce the association of more than one target (*4*–*11*). This modular rapid access to high molecular weight, amphiphilicity, and rotational flexibility can provide useful chemical probes and therapeutic leads for intracellular targets, as long as the resulting molecules remain cell-permeable.

Mechanisms to understand and predict the cell permeability of linked chemotypes, however, remain limited. Other medium-to-high molecular weight therapeutics such as natural products and synthetic macrocycles often

constitute highly tailored arrangements of polar and nonpolar functionality that allow switching between membrane-favored and aqueous-favored conformations to enable passive permeability through membranes (*12*). Additionally, cell-penetrating proteins/peptides commonly require appendage of highly charged moieties to enable electrostatic interactions with the plasma membrane and subsequent internalization (*13*–*15*). Studies involving the most rapidly expanding linked chemotype class—proteolysis-targeting chimeras (PROTACs) (*16*)—provide varying insights into the determinants of cell permeability (*17*–*22*). Despite their atypical properties, PROTACs and additional large molecules such as the dimeric immunophilin ligand rimiducid have shown in-cell activity robust enough to enter clinical trials (*16*, *23*).

Given this discrepancy between the favorable biological activity of many large, bivalent molecules and traditional concepts of passive permeability, we inferred that linked chemotypes might hijack cellular processes to assist with passage through the cell membrane. We selected as an example a bitopic inhibitor of MTOR, RapaLink-1 (*7*), whose molecular weight (1784 g/mol) falls well beyond common guidelines (≤ 500 g/mol) (*1*), and even beyond that of typical PROTACs (800 to 1200 g/mol) (table S1) (*18*). RapaLink-1's atypical structure—composed of the allosteric MTOR inhibitor rapamycin and the active site inhibitor sapanisertib linked by an 8-unit polyethylene glycol (PEG8) tether (Fig. 1A)—confers enhanced selectivity for MTOR complex 1 over MTOR complex 2 (*7*, *24*, *25*). The molecule is highly active in vivo, penetrates the blood-brain barrier, and serves as a prototype for the clinical candidate RMC-5552 (*7*, *24*–*27*), establishing itself as a drug-like compound that defies most traditional notions of drug-like structures. We hypothesized that cellular mechanisms assisting RapaLink-1's cytoplasmic entry could be identified by systematically perturbing genes that modulate the molecule's ability to reach and inhibit its intracellular target.

Complementary genome-scale chemical-genetic approaches identify IFITMs as regulators of RapaLink-1 cellular activity

We probed canonical protein-coding genes for cellular factors that determine RapaLink-1 uptake and sensitivity using a dCas9-based CRISPRi/a functional genomics platform (*28*, *29*). Gene expression inhibition and activation, through CRISPRi and CRISPRa, respectively, act as complementary approaches to map chemical-genetic interactions at the genome scale. In particular, genes displaying strong mirrored phenotypes (resistance upon knockdown and sensitivity upon overexpression) are likely to be directly involved in a small molecule's mechanism of action. This integrated approach to identifying physiologically relevant chemical-genetic interactions was proposed by Jost *et al.* and its utility has recently been reviewed (*30*, *31*). In addition to the bitopic inhibitor, we also assessed sapanisertib, rapamycin, and an unlinked control (a 1:1 mixture of sapanisertib and rapamycin) to distinguish chemical-genetic interactions specific to the linked chemotype (Fig. 1A).

Patient-derived chronic myeloid leukemia (CML) cells, K562, pre-engineered to express CRISPRi or CRISPRa machinery, were transduced with their respective genome-scale sgRNA libraries, selected with puromycin to remove nontransduced cells, and treated with DMSO, sapanisertib, rapamycin, sapanisertib + rapamycin, or RapaLink-1. The experiments were conducted with high-replicate reproducibility (fig. S1, A to D), and data from the genome-scale CRISPRi (data S1 and S2) and CRISPRa (data S3 and S4) screens were juxtaposed to highlight genes that displayed mirrored phenotypes (Fig. 1B). This arrangement distributes genes which functionally synergize with the inhibitor in the lower right (e.g., *FKBP12*, the required inhibitory complex partner of rapamycin) and those which antagonize the inhibitor in the upper left (e.g., *MTOR*, the direct target) (*30*, *31*). Chemical-genetic interactions with MTOR signaling components, particularly the Ragulator complex (*RRAGA*, *RRAGC*, and *LAMTOR1-5*) and nodes downstream of PI3K/AKT (*TSC1*, *TSC2*, and *RHEB*), were observed across multiple inhibitor conditions (fig. S2, A and B), consistent with known pathway relationships (*32*) and prior functional genomics studies (*33*, *34*).

A distinct set of chemical-genetic interactions were identified as top hits with RapaLink-1 and not with any of the nonlinked molecules tested, suggesting the involvement of a biological pathway that promotes the activity of the linked chemotype. The expression of members of a highly homologous gene family—interferon-induced transmembrane proteins (IFITMs) *IFITM1*, *IFITM2*, and *IFITM3* (*35*)—synergized with the activity of RapaLink-1 and not its nonlinked counterparts, sapanisertib and rapamycin (Fig. 1B). To validate this finding, we tested sgRNAs targeting *IFITM1-3* individually for transcriptional repression or activation (fig. S3A and table S2). CRISPRi-mediated knockdown of *IFITM1-3* was potent and selective (fig. S3B). CRISPRa-mediated overexpression was also potent although we observed variable cross activation between family members (fig. S3C), possibly due to concerted transcriptional regulation of these genes, which are adjacent to each other on chromosome 11 (fig. S3A) (*36*). We individually confirmed that top screen hits, including *FKBP12* and *IFITM1-3*, synergized with RapaLink-1 in a competitive growth assay; we also validated that the *IFITM1-3* chemical-genetic interaction was specific to the linked chemotype (fig. S3, D and E).

Seeking to generalize these observations beyond a single cell type, we employed an independent chemical-genetic approach correlating MTOR inhibitor sensitivity data with basal gene expression in diverse in vitro models (*37*–*39*). More than 500 cancer cell lines were assessed for sensitivity to sapanisertib, rapamycin, or RapaLink-1 (measured by area under the dose-response curve). These measurements were correlated with gene transcript abundance (measured by RNA sequencing) across the cell lines to identify predictive biomarkers for compound sensitivity or resistance. High expression of any of the three IFITM family members was strongly associated with enhanced RapaLink-1 sensitivity across 659 cell lines, and *IFITM2* was notably the single most associated sensitizing biomarker (negative correlation) for Rapalink-1 (Fig. 1C and data S5). This correlation was absent for sapanisertib and rapamycin (fig. S4, A to C, and data S5), recapitulating the CRISPRi/a screens. Together, our analysis of the CRISPRi/a screens and large-scale chemogenomic cell line profiling experiments suggested a role of IFITMs in promoting the activity of RapaLink-1 across diverse cell types and levels of IFITM expression.

IFITMs promote RapaLink-1 pharmacodynamic target engagement

To unmask potentially overlapping IFITM functions, we knocked down *IFITM1*, *IFITM2*, and *IFITM3* expression simultaneously by coexpressing three different targeting sgRNAs (table S2) (*40*). Although multigene knockdown was potent (fig. S5A), we did not observe baseline changes in cell viability (fig. S5B). *IFITM1-3* triple knockdown ablated MTOR inhibition by 3 nM RapaLink-1 in cells, as determined by intracellular markers of MTOR pathway signaling, phospho-$S6^{S235/236}$, phospho-$4EBP1^{T37/46}$, and phospho-AKT^{S473} (Fig. 1D),

[1]Department of Cellular and Molecular Pharmacology, University of California, San Francisco, San Francisco, CA 94158, USA. [2]Howard Hughes Medical Institute, University of California, San Francisco, San Francisco, CA 94158, USA. [3]Department of Pharmacological Sciences, Stony Brook University, Stony Brook, NY 11794-8651, USA. [4]Department of Chemistry, University of California, Berkeley, Berkeley, CA 94720, USA. [5]Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, San Francisco, CA 94158, USA. [6]Department of Urology, University of California, San Francisco, San Francisco, CA 94158, USA. [7]Center for Cancer Research, Massachusetts General Hospital Cancer Center, Charlestown, MA 02129, USA. [8]Department of Medicine, Harvard Medical School, Boston, MA 02115, USA. [9]Innovative Genomics Institute, University of California, San Francisco, San Francisco, CA 94158, USA. [10]Arc Institute, Palo Alto, CA 94304, USA.
***Corresponding author. Email: kevan.shokat@ucsf.edu (K.M.S.); luke@arcinstitute.org (L.A.G.)**

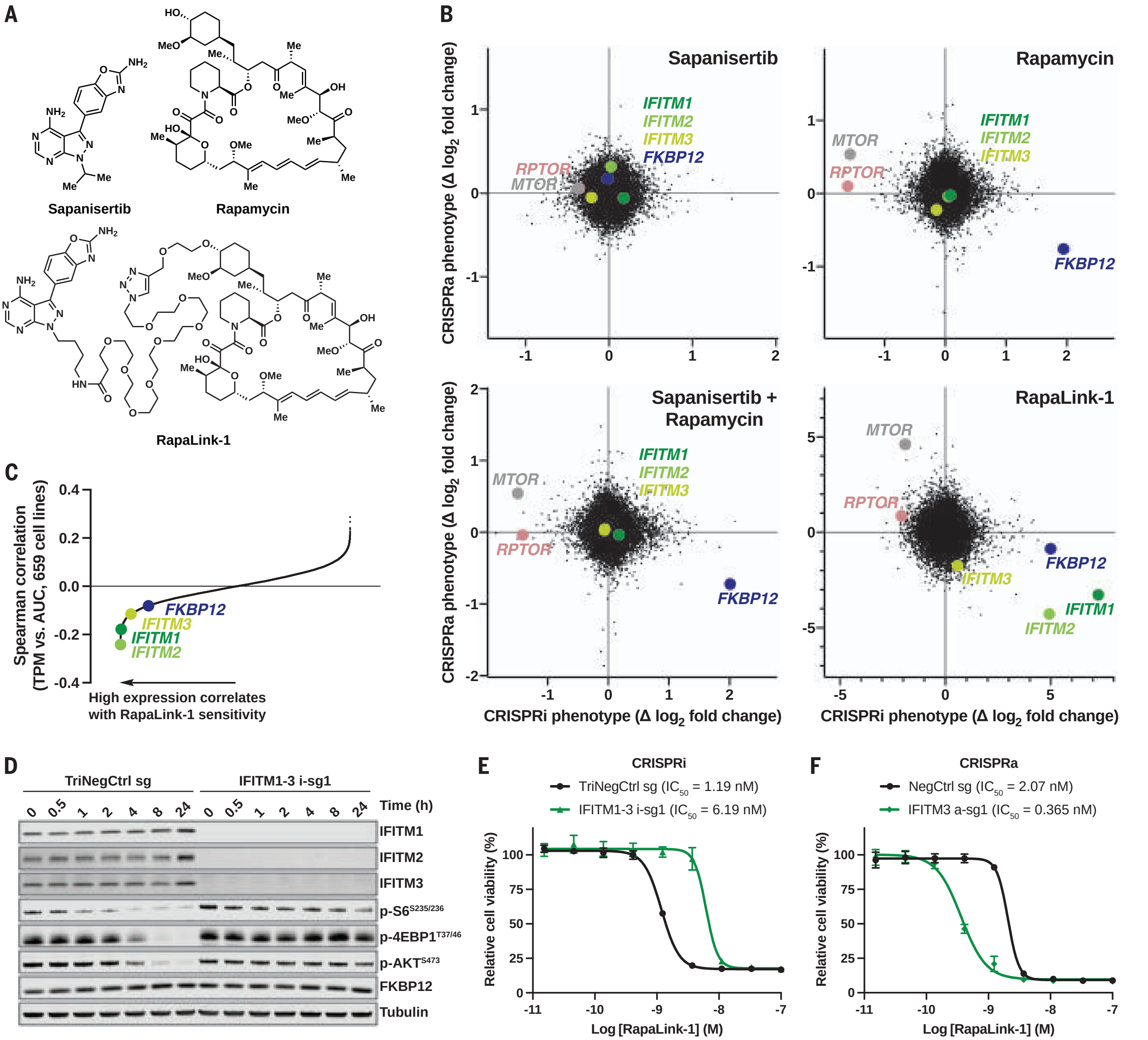


Fig. 1. IFITMs promote the cellular activity of a bitopic MTOR inhibitor. (**A**) Chemical structures of MTOR inhibitors. (**B**) Gene phenotypes from genome-scale CRISPRi and CRISPRa screens in K562 cells. Genes involved in MTOR complex 1 (*MTOR* and *RPTOR*), a requisite rapamycin inhibitory complex partner (*FKBP12*), and clade I IFITMs (*IFITM1*, *IFITM2*, and *IFITM3*) are highlighted. Data represent two biological replicates. (**C**) Spearman correlation coefficients between RapaLink-1 sensitivity, as measured by dose-response data, and transcript abundance, as measured by RNA sequencing (see also fig. S4). Dose-response data are expressed as area under the curve (AUC) and RNA sequencing data are expressed as transcripts per million (TPM). Genes are highlighted as in (B). (**D**) Immunoblots of K562 CRISPRi cells expressing sgRNAs treated with RapaLink-1 (3 nM) for the times indicated. [(**E**) and (**F**)] Viability of K562 CRISPRi (E) or CRISPRa (F) cells expressing sgRNAs treated with RapaLink-1. Data represent means of three biological replicates; error bars denote SD.

and conferred resistance to the linked molecule (Fig. 1E and fig. S5C). Overall, IFITM expression perturbation by CRISPRi and CRISPRa caused a combined 29.5-fold modulation in cellular potency of the molecule (Fig. 1, E and F). As observed previously (*7*), RapaLink-1 requires multiple hours to achieve maximal pharmacodynamic inhibition (Fig. 1D). This contrasts with the typical finding that small molecules reach their intracellular targets on a timescale of seconds to minutes (*41*) and may reflect how linked chemotypes exhibit distinct permeability characteristics from traditional drug-like molecules.

Neither of the nonlinked MTOR inhibitors tested demonstrated chemical-genetic interactions with *IFITM1-3* and thus IFITMs likely do not directly modulate MTOR signaling but instead cooperate with some aspect of RapaLink-1 not shared with the other inhibitors. Clade I IFITM family members, *IFITM1-3*, are closely related broad spectrum viral restriction factors that localize to the plasma and endolysosomal membranes (*42–45*). They are thought to perform their antiviral function in part by rendering local membrane characteristics at the viral-endosomal juncture unfavorable for viral entry (*46*), although in some cases viruses can also hijack IFITMs to facilitate entry and infection (*47*). In addition to their established immunologic function, clade I IFITMs are also reported to modulate an oncogenic phenotype (*48*), affect placenta formation (*49*),

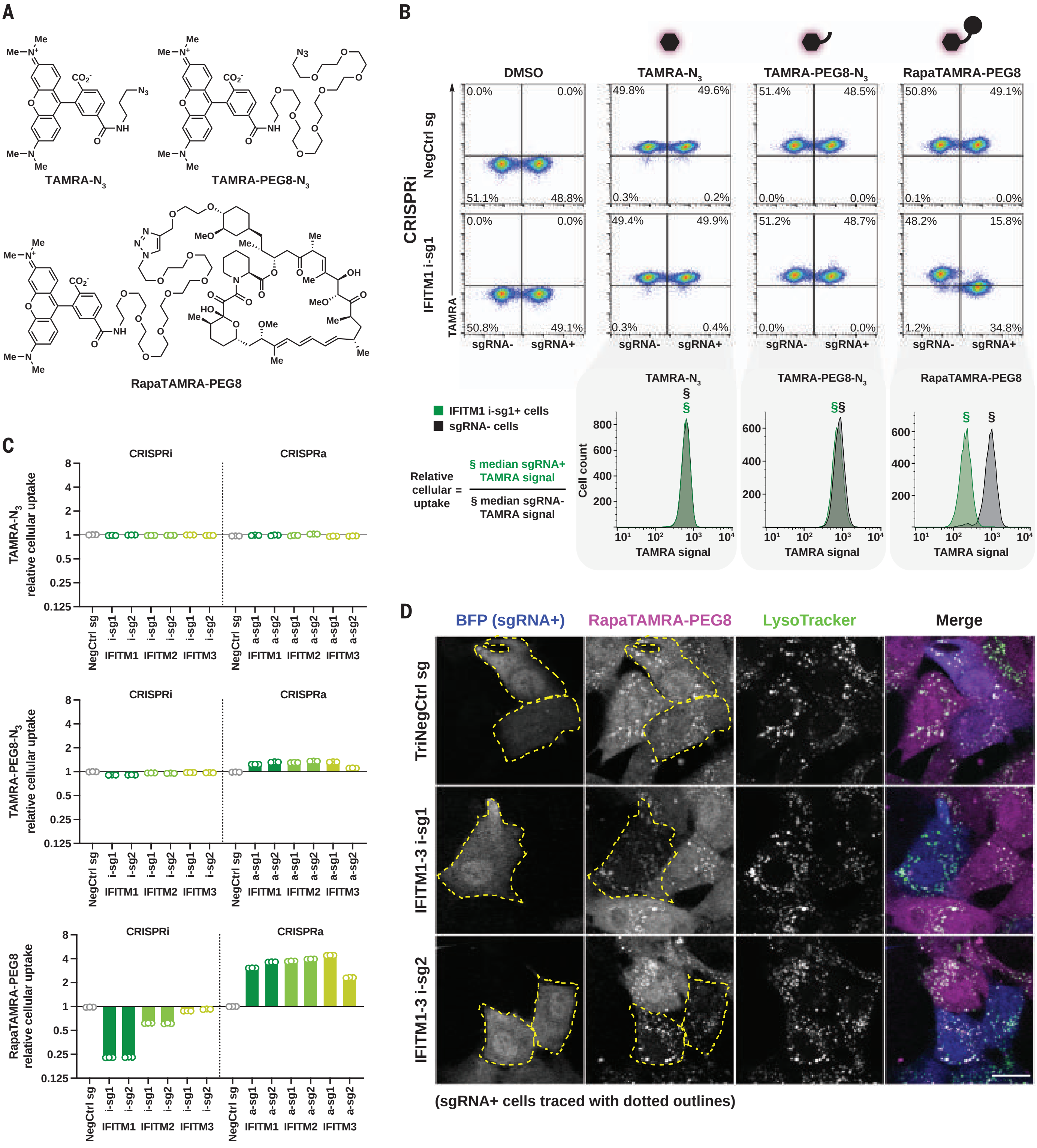


Fig. 2. IFITMs promote the cellular uptake of linked chemotypes. (**A**) Chemical structures of fluorescent RapaLink-1 analogs. (**B**) Measurement of fluorescent molecule uptake in K562 CRISPRi cells expressing sgRNAs (sgRNA+). Cells were incubated with TAMRA-N_3 (10 nM), TAMRA-PEG8-N_3 (1 μM), or RapaTAMRA-PEG8 (1 nM) for 24 hours. Uptake modulation by sgRNAs was quantified by internal normalization to nontransduced cells (sgRNA-) present within the mixture (i.e., relative cellular uptake). Data representative of three biological replicates. (**C**) Changes in uptake of fluorescent molecules by sgRNAs targeting *IFITM1-3* as in (B) and fig. S6A. Relative cellular uptake <1 indicates decreased uptake and >1 indicates increased uptake. Data represent means of three biological replicates. (**D**) Confocal microscopy images of RPE-1 CRISPRi cells expressing indicated sgRNAs (blue) and treated for 24 hours with RapaTAMRA-PEG8 (magenta) and LysoTracker (green). sgRNA+ cells are traced with dotted outlines (yellow) in left two columns for clarity. Scale bar denotes 20 μm.

and contribute to cellular homeostasis (*44*). In turn, Rapalink-1 might interact with IFITMs through a cellular pathway that promotes uptake of large molecules.

A fluorescent RapaLink-1 analog reveals a role for IFITMs in linked chemotype uptake

To explore our uptake hypothesis, we created a fluorescent analog of RapaLink-1 to directly observe the effect of IFITM expression on accumulation of the linked chemotype in live cells. This fluorescent molecule, RapaTAMRA-PEG8, was designed by replacing the adenosine triphosphate (ATP)-site binding element in RapaLink-1 with tetramethylrhodamine (TAMRA), resulting in a fluorescent derivative that closely mimics the physicochemical properties of the original molecule (Fig. 2A and table S1) (*27*). Analogs representing partial components of RapaTAMRA-PEG8, TAMRA-N_3, and TAMRA-PEG8-N_3, were additionally evaluated to assess whether the uptake pathway extended to generic compact-hydrophobic or linked-amphiphilic chemotypes respectively (Fig. 2A). We quantified accumulation of these molecules by flow cytometry using a quantitative live cell fluorescence uptake assay in which a mixture of transduced (sgRNA$^+$) and nontransduced (sgRNA$^-$) cells were equally exposed to compounds within the same well (Fig. 2B and fig. S6A). Changes in cellular uptake resulting from CRISPRi/a expression modulation by sgRNAs (fig. S3, A to C) again revealed a chemotype-specific IFITM dependency pattern (Fig. 2, B and C, and fig. S6A). Both linked chemotypes—TAMRA-PEG8-N_3 and RapaTAMRA-PEG8—demonstrated decreased uptake upon knockdown of *IFITM1-3* and increased uptake upon overexpression. The linker-less chemotype TAMRA-N_3, by contrast, exhibited no such chemical-genetic interactions. CRISPRi/a-induced uptake differences observed for RapaTAMRA-PEG8 correlated strongly with resistance and sensitivity phenotypes for RapaLink-1 (fig. S6B), suggesting a direct association between measured uptake and functional target inhibition. The observation that uptake of TAMRA-PEG8-N_3—a generic linked chemotype not specifically bound by any cellular protein—was also IFITM-assisted suggests that this uptake mechanism might also be used by other linked molecules.

Additionally, we assessed the role of IFITMs on the subcellular localization of RapaTAMRA-PEG8 by confocal microscopy in a human nontransformed cell line, RPE-1, pre-engineered to express CRISPRi machinery. Again a mixture of sgRNA$^+$ and sgRNA$^-$ cells were imaged in the same well following equal exposure to the fluorescent compound. *IFITM1-3* triple-knockdown significantly reduced the amount of RapaTAMRA-PEG8 entering the intracellular compartment (Fig. 2D and fig. S6C). A reduction in signal was also observed within the endolysosomal compartment (Fig. 2D and fig. S6C), suggesting that IFITMs, which localize to the plasma membrane as well as to endolysosomal membranes (*43–45*), may play a role in facilitating RapaTAMRA-PEG8 uptake through endocytic vesicles and into the intracellular space. Consistent with this, our functional genomics screens identified RapaLink-1–specific chemical-genetic interactions among endosomal (*ARF6*, *VPS26A*, *VPS29*, and *VPS35*) and sterol (*OSBP*, *GRAMD1A*, *INSIG1*, and *SCAP*) regulatory genes (fig. S7, A and B). This, in part, resembles the role IFITMs play as antiviral effectors in which biophysical interactions with incoming viral particles (*50*) and membrane sterols (*51*) may hinder or assist infection of target cells (*43–45*). Considering the large diversity of viruses IFITMs are described to interact with, we hypothesized that the uptake assistance afforded to RapaLink-1 and RapaTAMRA-PEG8 by IFITMs might also extend to other linked chemotypes with similar physicochemical properties.

DasatiLink-1 is an IFITM-assisted bitopic inhibitor of BCR-ABL1 with enhanced selectivity

To explore the generalizability of this IFITM-promoted cellular uptake mechanism, we designed, synthesized, and characterized a bitopic inhibitor that is, aside from being a linked molecule, compositionally unrelated to RapaLink-1. This inhibitor targets a different intracellular protein, BCR-ABL1, a fusion oncoprotein associated with CML and other leukemias (*52*). BCR-ABL1 harbors two well-defined small-molecule binding sites within its kinase domain (Fig. 3A): the ATP pocket (*53*), which is targeted by five clinically approved compounds (e.g., dasatinib) (*54*), and the myristoyl pocket (*55*), which is targeted by the clinical inhibitor asciminib (*56*). These sites can also be bound by the two classes of inhibitors simultaneously (*55–57*). The two pockets span a distance similar to that engaged by RapaLink-1 in MTOR (*7*), suggesting that a similar bitopic inhibitor linkage strategy could apply to BCR-ABL1. We devised a bitopic inhibitor of BCR-ABL1, DasatiLink-1, based on the linking of dasatinib and asciminib by a flexible tether whose length (41 heavy atoms) was close to that of RapaLink-1 (39 heavy atoms) (Fig. 3B).

We characterized the interaction between DasatiLink-1 and its target using in vitro biochemical assays. Treatment of purified BCR-ABL1 kinase domain with dasatinib or asciminib resulted in marked (>0.1 ppm) nuclear magnetic resonance (NMR) chemical shift differences in residues involved in binding to the monomeric inhibitors (fig. S8, A and B), consistent with previous reports (*55*, *56*, *58*). The NMR spectrum observed in the presence of Dasatilink-1 closely matched the spectrum observed with a mixture of the two nonlinked inhibitors (fig. S8, A and B), suggesting that the linked inhibitor simultaneously binds to both sites and that the tether does not prevent binding to either site.

We hypothesized that DasatiLink-1 might require an allosteric foothold to achieve high occupancy of the BCR-ABL1 kinase domain, and we tested this hypothesis using a pull-down assay for ATP site availability (*59*). We confirmed that the assay recapitulated a biochemical IC_{50} of <1 nM for dasatinib (*60*), which was unaffected by inclusion of 100-fold excess of the allosteric inhibitor asciminib (fig. S8C). However, addition of excess asciminib impaired the ability of DasatiLink-1 to occupy the ATP site, likely resulting from a loss of avidity following steric occlusion of the allosteric pocket (fig. S8C). This indicates that Dasatilink-1 relies on both the orthosteric and allosteric sites for binding, suggesting that it might also exhibit the enhanced selectivity often observed in bitopic inhibitors (*61*). Together, these biochemical data validate DasatiLink-1 as a bitopic inhibitor with physicochemical properties beyond standard drug design limits (*1–3*), and we reasoned that the molecule's linked composition might allow it to be assisted into the cell by IFITMs.

Returning to our K562 CRISPRi/a models, which are patient-derived BCR-ABL1 mutant CML cells, we characterized the effect of IFITM expression on the ability of DasatiLink-1 to inhibit intracellular BCR-ABL1 signaling. Similar to RapaLink-1 (Fig. 1, E and F), CRISPRi and CRISPRa perturbation of IFITM expression resulted in a combined 8.9-fold modulation of DasatiLink-1 cellular potency (Fig. 3, C and D). We also probed the capacity of DasatiLink-1 to engage intracellular BCR-ABL1 by measuring pharmacodynamic markers of inhibition. Consistent with an IFITM-assisted uptake mechanism, *IFITM1-3* triple knockdown reduced the ability of DasatiLink-1 to inhibit phospho-BCR-ABL1^{Y245} and phospho-STAT5^{Y694}, which are markers of BCR-ABL1 activity (Fig. 3E). The maximal inhibition observed for DasatiLink-1 in the TriNegCtrl sg conditions at 8 and 24 hours was not ever reached in the *IFITM1-3* triple knockdown conditions, likely due to lower intracellular concentrations of compounds resulting from decreased uptake. The inhibition kinetics we observed for the negative control treatment, requiring multiple hours for maximal inhibition at a nanomolar concentration (Fig. 3E), were also exhibited by RapaLink-1 (Fig. 1D).

We anticipated that DasatiLink-1, akin to RapaLink-1, was likely to be selective for its target as a result of its multivalent binding mechanism—only BCR-ABL1 kinase domain contains binding sites for both of its linked components. We assessed DasatiLink-1's kinome-wide selectivity in live cells using a promiscuous kinase occupancy probe, XO44 (*62*), with which kinase active site occupancy can

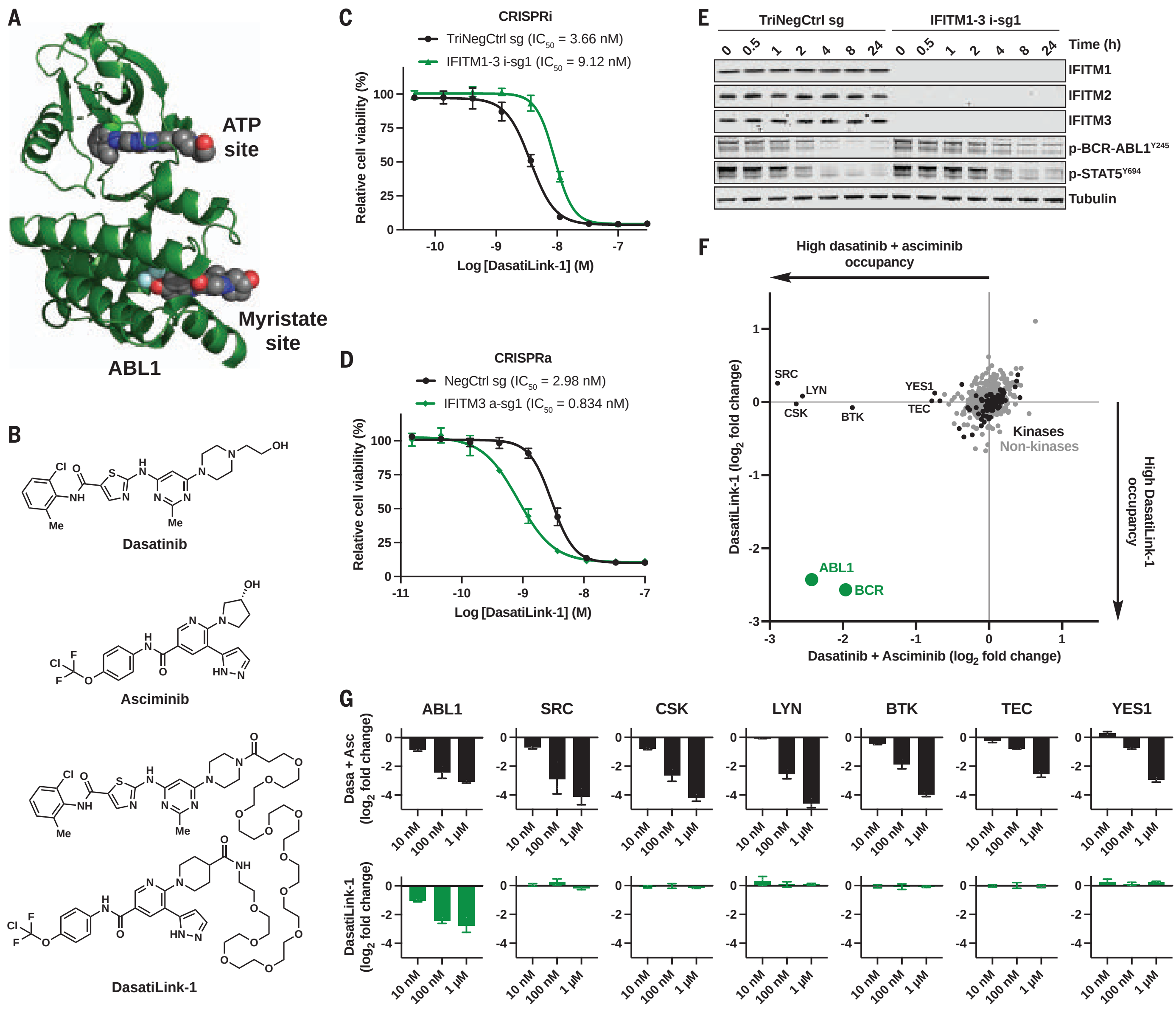


Fig. 3. Design and characterization of an IFITM-assisted bitopic BCR-ABL1 inhibitor. (**A**) Molecular model of ABL1 kinase domain. The model was constructed by aligning two crystal structures: one bound to dasatinib (PDB, 2GQG) and one bound to asciminib (PDB, 5MO4). (**B**) Chemical structures of BCR-ABL1 inhibitors. [(**C**) and (**D**)] Viability of K562 CRISPRi (C) or CRISPRa (D) cells expressing sgRNAs treated with DasatiLink-1. Data represent means of three biological replicates; error bars denote SD. (**E**) Immunoblots of K562 CRISPRi cells expressing sgRNAs treated with DasatiLink-1 (5 nM) for the times indicated (**F**) In-cell kinase occupancy profiling of DasatiLink-1 and an unlinked control (a 1:1 mixture of dasatinib and asciminib) at equimolar concentration (100 nM). Data represent means of three biological replicates. (**G**) As in (F) for kinases occupied following 10 nM, 100 nM, and 1 μM inhibitor treatments; error bars denote SD.

be determined through competitive activity-based protein profiling (*63*). In contrast to an unlinked control (a 1:1 mixture of dasatinib and asciminib) at equimolar concentration, which competed with XO44 for labeling of numerous known dasatinib targets (*62*), pretreatment with DasatiLink-1 resulted in observable intracellular occupancy of only ABL1 (Fig. 3F and data S6). This single kinase specificity extended over a 100-fold concentration range up to 1 μM, the same range over which the unlinked control demonstrated dose-responsive occupancy of numerous off targets (Fig. 3G and data S6). These data suggest that target selectivity can be conferred by two-site binding, analogous to RapaLink-1's selectivity for MTOR complex 1 (*7*, *24*, *25*).

BisRoc-1 analogs reveal linker length dependency of IFITM-assisted cellular uptake

To further examine the breadth of linked chemotypes that might be assisted by IFITMs, we designed, synthesized, and characterized a new linked molecular glue inhibitor based on the natural product rocaglamide. Rocaglamide clamps the EIF4A1 helicase to 5′ untranslated regions (UTRs) of target mRNAs to inhibit the translation of downstream sequences (*64*). The crystal structure of the complex of rocaglamide, EIF4A1, and polypurine RNA (*65*) revealed that the molecule's amide points toward free solvent, near a symmetry mate (fig. S9A). We reasoned that dimerization of rocaglamide through its amide position could be a chemically tractable means to simultaneously engage two proximal EIF4A1-RNA complexes within the cell. We designed a molecule, BisRoc-1 (fig. S9B), that links two rocaglamide monomers together with a linker length (35 heavy

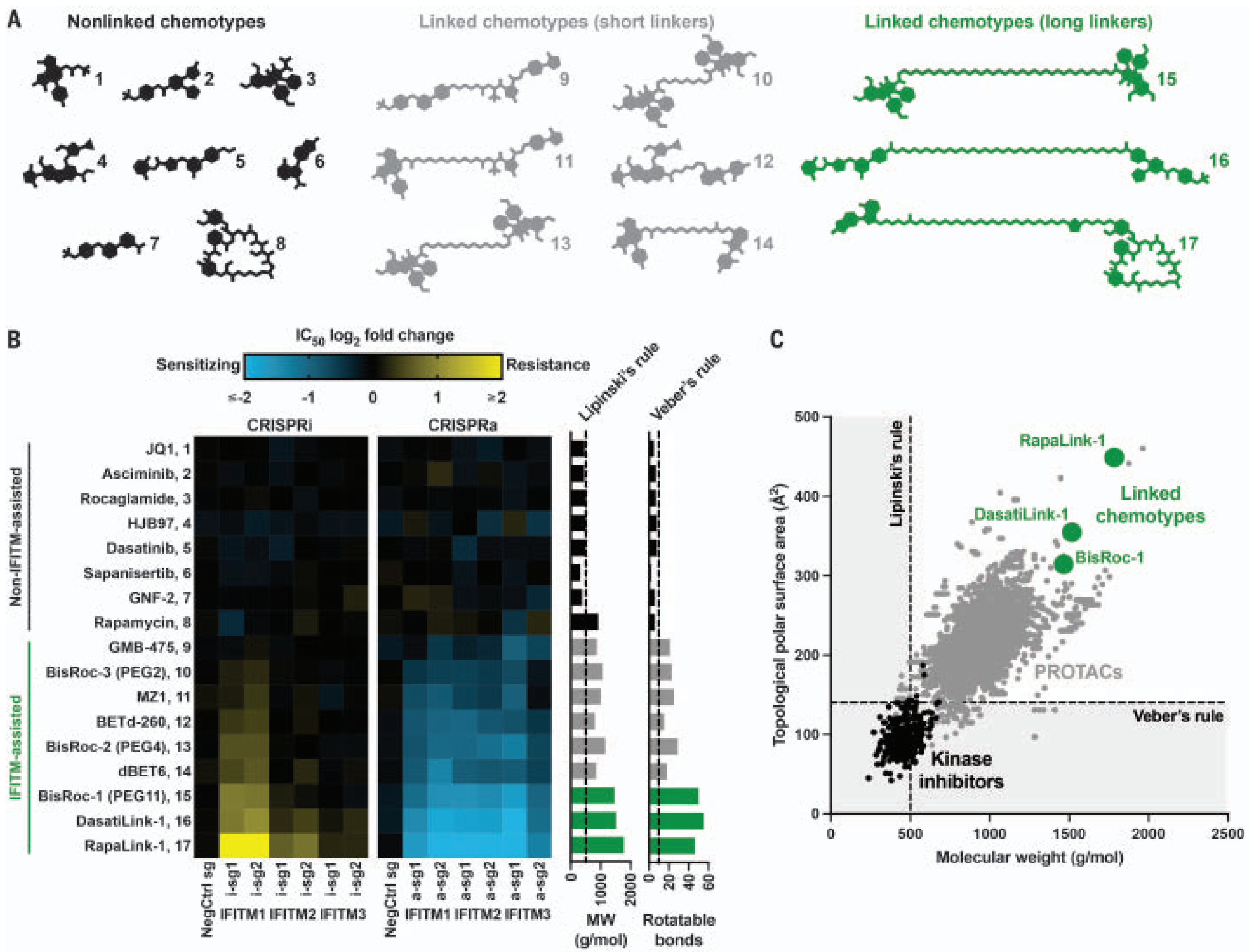


Fig. 4. IFITMs assist the cellular activity of diverse linked chemotypes. (**A**) Heavy atom skeletons of compounds assessed for IFITM assistance (see also fig. S10 for chemical structures). Compounds were categorized as nonlinked chemotypes (compounds 1 to 8, black), linked chemotypes with short linkers (compounds 9 to 14, gray), or linked chemotypes with long linkers (compounds 15 to 17, green). (**B**) Chemical-genetic interaction map of inhibitors in (A) with *IFITM1*, *IFTM2*, and *IFITM3*. Potency, as measured by dose-response IC_{50} in a cell viability assay (see also Fig. 1F, Fig. 3D, or fig. S9D for example source data), was normalized to that of non-sgRNA–expressing K562 CRISPRi or CRISPRa cells. Physicochemical properties, including molecular weight (MW) and number of rotatable bonds, with their respective traditional thresholds for drug-likeness are indicated (right). Data represent means of three biological replicates. (**C**) Map of chemical space populated by 304 kinase inhibitors in clinical development (black), 3270 PROTACs reported in the literature (gray), and 3 linked chemotypes described herein (green). Boundaries represent traditional guidelines for drug-likeness.

atoms) exceeding the distance separating two rocaglamide binding sites in the crystal structure (fig. S9A). Similar to RapaLink-1 and DasatiLink-1, CRISPRi and CRISPRa perturbation of IFITM expression resulted in a combined 6.2-fold modulation in cellular activity of BisRoc-1 (fig. S9, C and D). Additionally, we evaluated the relationship between linker length and IFITM assistance by examining an analog series consisting of BisRoc-1 (PEG11), BisRoc-2 (PEG4), BisRoc-3 (PEG2), and rocaglamide (no linker) (fig. S9E). We treated our K562 CRISPRi and CRISPRa cells with these inhibitors and evaluated differences in potency resulting from IFITM expression modulation, as measured by a half-maximal inhibitory concentration (IC_{50}) shift in a cell viability assay (fig. S9, E and F, and data S7). This revealed a pattern in which longer linker lengths correlated with greater IFITM assistance. Combined, these data suggest the general feasibility of retaining cell permeability despite increased pharmacophore size, polarity, and flexibility in the context of linked chemotypes described herein.

An expanded chemical space for cell-permeable molecules

Given the ubiquitous presence of IFITMs in cells, we hypothesized that the cellular uptake of other linked inhibitors in the literature could also be assisted by IFITMs. Although generally not as large as the linked chemotypes described above, PROTACs are likewise composed of two chemical entities covalently attached by a flexible tether (*16*). Thus, we included four PROTACs (GMB-475, MZ1, BETd-260, and dBET6) and their nonlinked parent inhibitors in an expanded survey of chemical-genetic interactions with IFITMs (Fig. 4A, fig. S10, and table S1). These compounds were evaluated in the same IFITM dependency analysis as the BisRoc linker series (Fig. 4B and data S7). Using RapaLink-1 as a chemical benchmark, we observed that *IFITM1*, *IFITM2*, and *IFITM3* overexpression sensitized cells to linked chemotypes (Fig. 4B; compounds 9 to 17). The inverse finding—resistance to linked

chemotypes—resulted from gene knockdown (Fig. 4B). The trend observed in the BisRoc series was corroborated across the nine bivalent molecules tested (Fig. 4B): the magnitudes of chemical-genetic interactions correlated with linker length which is reflected in inhibitor size (molecular weight) and flexibility (number of rotatable bonds). Linked chemotypes with long linkers were more IFITM-assisted than linked chemotypes with short linkers, and nonlinked chemotypes (Fig. 4B; compounds 1 to 8) were not observed to be assisted by IFITMs (Fig. 4B). Despite their cellular activities, the physicochemical properties of the linked chemotypes largely violate Lipinski's (*1*) and Veber's (*2*) classic guidelines (Fig. 4, A to C, and table S1), raising the need for a revised drug design framework that considers IFITM-assisted uptake and other cellular import processes.

Discussion

Through a combination of functional genomics and chemical methods, we uncovered an endogenous pathway involving IFITMs that, in our data, promotes the cellular uptake of diverse linked chemotypes. With the clinical advancement of a dimeric immunophilin ligand (*23*), PROTACs (*16*), and a RapaLink-1 derivative (*26*), the notion of "drug-like" is continually being revised. As evidence, the chemical space (*66*) populated by an ever-expanding set of linked preclinical compounds in the literature ventures beyond that occupied by lead inhibitors developed under traditional guidelines (Fig. 4C) (*1–3*). Here, we identify IFITM-assisted cellular uptake as one of the mechanisms by which linked inhibitors are able to break previously established rules surrounding drug likeness. We anticipate that our findings will inform the uptake optimization of emerging classes of bivalent molecules (PROTACs, Syn-TEFs, RIBOTACs, PHICS, DUBTACs, and others) (*4–11*) and enable the design of cell-permeable therapeutics that bridge distal binding sites on solitary targets or multitarget complexes.

REFERENCES AND NOTES

1. C. A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeney, *Adv. Drug Deliv. Rev.* **46**, 3–26 (2001).
2. D. F. Veber *et al.*, *J. Med. Chem.* **45**, 2615–2623 (2002).
3. A. K. Ghose, V. N. Viswanadhan, J. J. Wendoloski, *J. Comb. Chem.* **1**, 55–68 (1999).
4. M. Erez, A. E. Takemori, P. S. Portoghese, *J. Med. Chem.* **25**, 847–849 (1982).
5. D. M. Spencer, T. J. Wandless, S. L. Schreiber, G. R. Crabtree, *Science* **262**, 1019–1024 (1993).
6. K. M. Sakamoto *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 8554–8559 (2001).
7. V. S. Rodrik-Outmezguine *et al.*, *Nature* **534**, 272–276 (2016).
8. G. S. Erwin *et al.*, *Science* **358**, 1617–1622 (2017).
9. M. G. Costales, Y. Matsumoto, S. P. Velagapudi, M. D. Disney, *J. Am. Chem. Soc.* **140**, 6741–6744 (2018).
10. S. U. Siriwardena *et al.*, *J. Am. Chem. Soc.* **142**, 14052–14057 (2020).
11. N. J. Henning *et al.*, *Nat. Chem. Biol.* **18**, 412–421 (2022).
12. B. C. Doak, B. Over, F. Giordanetto, J. Kihlberg, *Chem. Biol.* **21**, 1115–1142 (2014).
13. M. Green, P. M. Loewenstein, *Cell* **55**, 1179–1188 (1988).
14. A. D. Frankel, C. O. Pabo, *Cell* **55**, 1189–1193 (1988).
15. E. L. Snyder, S. F. Dowdy, *Pharm. Res.* **21**, 389–393 (2004).
16. M. Békés, D. R. Langley, C. M. Crews, *Nat. Rev. Drug Discov.* **21**, 181–200 (2022).
17. C. A. Foley, F. Potjewyd, K. N. Lamb, L. I. James, S. V. Frye, *ACS Chem. Biol.* **15**, 290–295 (2020).
18. C. Cantrill *et al.*, *Drug Discov. Today* **25**, 969–982 (2020).
19. D. E. Scott *et al.*, *ACS Med. Chem. Lett.* **11**, 1539–1547 (2020).
20. G. Ermondi, M. Vallaro, G. Caron, *Drug Discov. Today* **25**, 1585–1591 (2020).
21. V. G. Klein *et al.*, *ACS Med. Chem. Lett.* **11**, 1732–1738 (2020).
22. Y. Atilaw *et al.*, *ACS Med. Chem. Lett.* **12**, 107–114 (2020).
23. A. D. Stasi *et al.*, *N. Engl. J. Med.* **365**, 1673–1683 (2011).
24. Q. Fan *et al.*, *Cancer Cell* **31**, 424–435 (2017).
25. B. J. Lee *et al.*, *Nat. Chem. Biol.* **17**, 1065–1074 (2021).
26. H. A. Burris III *et al.*, *J. Clin. Oncol.* **40**, 3098 (2022).
27. Z. Zhang *et al.*, *Nature* **609**, 822–828 (2022).
28. L. A. Gilbert *et al.*, *Cell* **159**, 647–661 (2014).
29. M. A. Horlbeck *et al.*, *eLife* **5**, e19760 (2016).
30. M. Jost *et al.*, *Mol. Cell* **68**, 210–223.e6 (2017).
31. M. Jost, J. S. Weissman, *ACS Chem. Biol.* **13**, 366–375 (2018).
32. G. Y. Liu, D. M. Sabatini, *Nat. Rev. Mol. Cell Biol.* **21**, 183–203 (2020).
33. E. De Zan *et al.*, *Sci. Signal.* **13**, eaba5665 (2020).
34. K. J. Condon *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2022120118 (2021).
35. Z. Zhang, J. Liu, M. Li, H. Yang, C. Zhang, *PLOS ONE* **7**, e49265 (2012).
36. P. Li *et al.*, *Biochim. Biophys. Acta Gene Regul. Mech.* **1860**, 885–893 (2017).
37. M. J. Garnett *et al.*, *Nature* **483**, 570–575 (2012).
38. M. G. Rees *et al.*, *Nat. Chem. Biol.* **12**, 109–116 (2016).
39. F. Iorio *et al.*, *Cell* **166**, 740–754 (2016).
40. B. Adamson *et al.*, *Cell* **167**, 1867–1882.e21 (2016).
41. B. R. Stockwell, *Nat. Rev. Genet.* **1**, 116–125 (2000).
42. A. L. Brass *et al.*, *Cell* **139**, 1243–1254 (2009).
43. C. C. Bailey, G. Zhong, I.-C. Huang, M. Farzan, *Annu. Rev. Virol.* **1**, 261–283 (2014).
44. G. Shi, O. Schwartz, A. A. Compton, *Retrovirology* **14**, 53 (2017).
45. X. Zhao, J. Li, C. A. Winkler, P. An, J.-T. Guo, *Front. Microbiol.* **9**, 3228 (2019).
46. K. Li *et al.*, *PLOS Pathog.* **9**, e1003124 (2013).
47. X. Zhao *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 6756–6761 (2014).
48. J. Lee *et al.*, *Nature* **588**, 491–497 (2020).
49. J. Buchrieser *et al.*, *Science* **365**, 176–180 (2019).
50. J. S. Spence *et al.*, *Nat. Chem. Biol.* **15**, 259–268 (2019).
51. T. Das *et al.*, *ACS Chem. Biol.* **17**, 2109–2120 (2022).
52. R. Kurzrock, J. U. Gutterman, M. Talpaz, M. Talpaz, *N. Engl. J. Med.* **319**, 990–998 (1988).
53. T. Schindler *et al.*, *Science* **289**, 1938–1942 (2000).
54. T. P. Braun, C. A. Eide, B. J. Druker, *Cancer Cell* **37**, 530–542 (2020).
55. J. Zhang *et al.*, *Nature* **463**, 501–506 (2010).
56. A. A. Wylie *et al.*, *Nature* **543**, 733–737 (2017).
57. R. E. Iacob, J. Zhang, N. S. Gray, J. R. Engen, *PLOS ONE* **6**, e15929 (2011).
58. N. Vajpai *et al.*, *J. Biol. Chem.* **283**, 18292–18302 (2008).
59. M. A. Fabian *et al.*, *Nat. Biotechnol.* **23**, 329–336 (2005).
60. J. Das *et al.*, *J. Med. Chem.* **49**, 6819–6832 (2006).
61. C. M. Gower, M. E. K. Chang, D. J. Maly, *Crit. Rev. Biochem. Mol. Biol.* **49**, 102–115 (2014).
62. Q. Zhao *et al.*, *J. Am. Chem. Soc.* **139**, 680–685 (2017).
63. B. F. Cravatt, A. T. Wright, J. W. Kozarich, *Annu. Rev. Biochem.* **77**, 383–414 (2008).
64. S. Iwasaki, S. N. Floor, N. T. Ingolia, *Nature* **534**, 558–561 (2016).
65. S. Iwasaki *et al.*, *Mol. Cell* **73**, 738–748.e9 (2019).
66. C. M. Dobson, *Nature* **432**, 824–828 (2004).
67. D. R. Wassarman, dwassarman/cellpanelr, v0.0.0.9001, Zenodo (2022).
68. M. A. Horlbeck, X. Xiong, M. Jost, mhorlbeck/ScreenProcessing, v0.1, Zenodo (2022).

ACKNOWLEDGMENTS

Sequencing was performed at the UCSF CAT, supported by UCSF PBBR, RRP IMIA, and NIH 1S10OD028511-01 grants. We thank T. S. Wu for computational advice, S. Kim and K. Herrington at the UCSF Nikon Imaging Center for their microscopy assistance and expertise, D. M. Peacock for assistance acquiring NMR characterization of compounds synthesized herein, and J. W. Stevenson for proofreading. **Funding:** This work was supported by NIH grant F30CA239476 (to K.L.); NIH grant F31CA243439 (to D.R.W.); Damon Runyon Cancer Research Foundation fellowship DRG-2281-17 (to Z.Z.); Wellcome Trust grant 102696 (to C.H.B.); NIH grant R35GM119437 (to M.A.S.); Ono Pharma Foundation and Pfizer (to J.T.); NIH New Innovator Award DP2-CA239597, a Pew-Stewart Scholars for Cancer Research award, and the Goldberg-Benioff Endowed Professorship in Prostate Cancer Translational Biology (to L.A.G.); the Howard Hughes Medical Institute, the Samuel Waxman Cancer Research Foundation, and NIH grants 1R01CA221969 and 1U19 AI171110-01 (to K.M.S.). **Author contributions:** K.L., L.A.G., and K.M.S. were responsible for the conception, design, and interpretation of the experiments and wrote the manuscript. K.L. performed CRISPRi/a experiments, cellular uptake assays, and chemical-genetic interaction mapping. K.L. designed and synthesized DasatiLink and BisRoc molecules. D.R.W., R.K.E., and P.G. performed sensitivity-expression correlation experiments. T.Y. performed chemoproteomic kinase target engagement assays. Y.P. performed protein NMR spectroscopy. Z.Z. designed and synthesized TAMRA molecules. T.A.O. performed fluorescent microscopy experiments. M.K.M. performed biological validation of BisRoc experiments. C.H.B., M.A.S., J.T., L.A.G., and K.M.S. supervised experiments and interpreted data. All authors edited the manuscript. **Competing interests:** K.L. and K.M.S. have filed a patent application covering DasatiLink analogs. K.L., M.K.M., and K.M.S. have filed a patent application covering BisRoc analogs. Kin of K.L. hold stock in and are employed by Pharmaron. C.H.B. is currently an employee of Novartis. J.T. is a founder of Global Blood Therapeutics, Principia Biopharma, Kezar Life Sciences, Cedilla Therapeutics, and Terremoto Biosciences, and is a scientific advisor to Entos. L.A.G. has filed patent applications related to CRISPRi/a screening. K.M.S. has filed patent applications covering RapaLink analogs which are licensed to Revolution Medicines. K.M.S. receives stock and cash compensation from Revolution Medicines. **Data and materials availability:** All data are available in the manuscript or the supplementary materials. Scripts implementing analyses are available at https://github.com/dwassarman, https://github.com/mhorlbeck, and Zenodo (*67*, *68*). Materials are available upon request to the corresponding authors with a signed material transfer agreement. This article is subject to HHMI's Open Access to Publications policy. HHMI lab heads have previously granted a nonexclusive CC BY 4.0 license to the public and a sublicensable license to HHMI in their research articles. Pursuant to those licenses, the author-accepted manuscript of this article can be made freely available under a CC BY 4.0 license immediately upon publication.

License information: Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.sciencemag.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abl5829
Chemical Synthesis
Materials and Methods
Figs. S1 to S10
Tables S1 and S2
References (*69–77*)
MDAR Reproducibility Checklist
Data S1 to S7

Submitted 26 July 2021; resubmitted 5 May 2022
Accepted 19 October 2022
10.1126/science.abl5829

MARTIAN GEOLOGY

Aqueous alteration processes in Jezero crater, Mars—implications for organic geochemistry

Eva L. Scheller[1,2]*†, Joseph Razzell Hollis[3,4]†, Emily L. Cardarelli[3], Andrew Steele[5], Luther W. Beegle[3], Rohit Bhartia[6], Pamela Conrad[5], Kyle Uckert[3], Sunanda Sharma[3], Bethany L. Ehlmann[1,3], William J. Abbey[3], Sanford A. Asher[7], Kathleen C. Benison[8], Eve L. Berger[9,10,11], Olivier Beyssac[12], Benjamin L. Bleefeld[13], Tanja Bosak[2], Adrian J. Brown[14], Aaron S. Burton[11], Sergei V. Bykov[7], Ed Cloutis[15], Alberto G. Fairén[16,17], Lauren DeFlores[3], Kenneth A. Farley[1], Deidra M. Fey[13], Teresa Fornaro[18], Allison C. Fox[11], Marc Fries[11], Keyron Hickman-Lewis[19,20], William F. Hug[6], Joshua E. Huggett[13], Samara Imbeah[13], Ryan S. Jakubek[11], Linda C. Kah[21], Peter Kelemen[22], Megan R. Kennedy[13], Tanya Kizovski[23], Carina Lee[24], Yang Liu[3], Lucia Mandon[25], Francis M. McCubbin[11], Kelsey R. Moore[3], Brian E. Nixon[13], Jorge I. Núñez[26], Carolina Rodriguez Sanchez-Vahamonde[13], Ryan D. Roppel[7], Mitchell Schulte[27], Mark A. Sephton[28], Shiv K. Sharma[29], Sandra Siljeström[30], Svetlana Shkolyar[31,32], David L. Shuster[33], Justin I. Simon[11], Rebecca J. Smith[34], Kathryn M. Stack[3], Kim Steadman[3], Benjamin P. Weiss[2], Alyssa Werynski[13], Amy J. Williams[35], Roger C. Wiens[36,37], Kenneth H. Williford[3,38], Kathrine Winchell[13], Brittan Wogsland[21], Anastasia Yanchilina[39], Rachel Yingling[13], Maria-Paz Zorzano[16]

The Perseverance rover landed in Jezero crater, Mars, in February 2021. We used the Scanning Habitable Environments with Raman and Luminescence for Organics and Chemicals (SHERLOC) instrument to perform deep-ultraviolet Raman and fluorescence spectroscopy of three rocks within the crater. We identify evidence for two distinct ancient aqueous environments at different times. Reactions with liquid water formed carbonates in an olivine-rich igneous rock. A sulfate-perchlorate mixture is present in the rocks, which probably formed by later modifications of the rocks by brine. Fluorescence signatures consistent with aromatic organic compounds occur throughout these rocks and are preserved in minerals related to both aqueous environments.

The Perseverance rover landed in Jezero crater, Mars, to investigate the geology of the crater, identify habitable environments, assess whether life ever existed on Mars, and collect samples for potential return to Earth (*1*). Jezero hosted an open-basin lake during the Noachian era (~3.7 billion years ago) (*1*, *2*), contains geologic units associated with the largest carbonate deposit identified on Mars (*2*–*4*), and contains a well-preserved delta with clay and carbonate-bearing sediments, which might contain organics (*1*). Organics have previously been de- tected on Mars (*5*, *6*).

We investigated the spatial and mineralogical context of organics in Jezero crater using the rover's Scanning Habitable Environments with Raman and Luminescence for Organics and Chemicals (SHERLOC) instrument, a deep-ultraviolet fluorescence and Raman spectrometer capable of mapping organic and mineral composition with a spatial resolution of 100 μm (*7*). Complementary elemental chemistry analyses were performed using the Planetary Instrument for X-ray Lithochemistry (PIXL) (*8*–*11*) and SuperCam instruments (*9*).

We identify organics and aqueously formed minerals at Jezero crater in three rock targets (*8*) analyzed during the first 208 martian days of the mission (Fig. 1) located in two different geological units within the floor of Jezero crater (*9*, *12*). The Garde target is from the altered ultramafic Séítah formation (Fm), orbitally mapped as the Crater Floor Fractured 1 unit (CF-f1) (Fig. 1) (*9*, *12*). The Guillaumes and Bellegarde targets are from the overlying, and therefore younger, basaltic Máaz Fm, orbitally mapped as the ~2.3-billion- to 2.6-billion-year-old (*13*) Crater Floor Fractured Rough unit (CF-fr) (*9*, *12*). The Perseverance rover drilled four rock samples from the Séítah Fm. Two drilled rock samples were obtained from the Bellegarde rock, whereas the Guillaumes drilled rock sample attempt, Roubion, failed (*12*). These six rock samples are planned to be returned to Earth.

All three Raman spectral scans (*8*) from Garde exhibit strong peaks at Raman shifts between 1080 and 1090 cm^{-1} (investigated in 38 separate scan points) that are attributed to carbonate [spectrum 1 and regions of interest (ROIs) 1 to 4 in Fig. 2H] and peaks with a peak position range of 820 to 840 cm^{-1} (n = 60) that are attributed to olivine (ROIs 1 and 4 in Fig. 2H) (*8*, *13*, *14*). Olivines were found to be more Fe-rich than laboratory measured olivines, with fosterite numbers [defined as $Mg/(Mg + Fe^{2+}) \times 100$] of 80 to 90 (*13*), whereas carbonates are likely mixed Fe- and Mg-species based on 1080- to 1087-cm^{-1} peak positions (*8*) and Ca-dominated species are excluded based on PIXL data (*11*). These spectral detections were overlaid on Wide-Angle Topographic Sensor for Operations and eNgineering (WATSON) camera images to compare spectral positions with textures (*8*). Olivine and carbonate are associated with micrometer- to millimeter-sized light-toned tan, reddish-brown, and dark-toned subangular grains as well as light-toned intergranular spaces (Fig. 2, B and E). Spectral features of olivines and carbonates often co-occur in a single spectrum; however, there are also areas where either olivine or carbonate occur independently. Spectral observations of a weak, broad Raman peak centered at ~1060 cm^{-1} (full width at half maximum ~200 cm^{-1}) could indicate a disordered phase consistent with amorphous silicates, which is often difficult

[1]Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA, USA. [2]Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA. [3]NASA Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA. [4]The Natural History Museum, London, UK. [5]Earth and Planets Laboratory, Carnegie Institution for Science, Washington, DC, USA. [6]Photon Systems Incorporated, Covina, CA, USA. [7]Department of Chemistry, University of Pittsburgh, Pittsburgh, PA, USA. [8]Department of Geology and Geography, West Virginia University, Morgantown, WV, USA. [9]Texas State University, San Marcos, TX, USA. [10]Jacobs Johnson Space Center Engineering, Technology and Science Contract, Houston, TX, USA. [11]NASA Johnson Space Center, Houston, TX, USA. [12]Institut de Minéralogie, de Physique des Matériaux et de Cosmochimie, Centre National de la Recherche Scientifique, Sorbonne Université, Muséum National d'Histoire Naturelle, 75005 Paris, France. [13]Malin Space Science Systems, San Diego, CA, USA. [14]Plancus Research, Severna Park, MD, USA. [15]Geography, The University of Winnipeg, Winnipeg, MB, Canada. [16]Centro de Astrobiología, Consejo Superior de Investigaciones Cientificas–Instituto Nacional de Tecnica Aeroespacial, Madrid, Spain. [17]Department of Astronomy, Cornell University, Ithaca, NY, USA. [18]Astrophysical Observatory of Arcetri, Istituto Nazionale di Astrofisica, Florence, Italy. [19]Department of Earth Sciences, The Natural History Museum, London, UK. [20]Dipartimento di Scienze Biologiche, Geologiche e Ambientali, Università di Bologna, Bologna, Italy. [21]Department of Earth and Planetary Sciences, University of Tennessee, Knoxville, TN, USA. [22]Lamont Doherty Earth Observatory, Columbia University, Palisades, NY, USA. [23]Department of Earth Sciences, Brock University, St. Catharines, ON L2S 3A1, Canada. [24]Lunar and Planetary Institute, Universities Space Research Association, Houston, TX, USA. [25]Laboratoire d'Etudes Spatiales et d'Instrumentation en Astrophysique, Observatoire de Paris, Centre National de la Recherche Scientifique, Sorbonne Université, Université Paris Diderot, 92195 Meudon, France. [26]Johns Hopkins University Applied Physics Laboratory, Laurel, MD, USA. [27]Mars Exploration Program, NASA Headquarters, Washington, DC, USA. [28]Earth Science and Engineering, South Kensington Campus, Imperial College London, SW7 2AZ London, UK. [29]Hawaii Institute of Geophysics and Planetology, University of Hawaii at Manoa, Honolulu, HI, USA. [30]Research Institutes of Sweden, Stockholm, Sweden. [31]Department of Astronomy, University of Maryland, College Park, MD, USA. [32]NASA Goddard Space Flight Center, Greenbelt, MD, USA. [33]Earth and Planetary Science, University of California Berkeley, Berkeley, CA, USA. [34]Department of Geosciences, Stony Brook University, Stony Brook, NY, USA. [35]Department of Geological Sciences, University of Florida, Gainesville, FL, USA. [36]Los Alamos National Laboratory, Los Alamos, NM, USA. [37]Earth, Atmospheric, and Planetary Sciences, Purdue University, West Lafayette, IN, USA. [38]Blue Marble Space Institute of Science, Seattle, WA, USA. [39]Impossible Sensing, St. Louis, MO, USA.
***Corresponding author. Email: eschelle@mit.edu**
†These authors contributed equally to this work.

to detect given the low band intensity (Fig. 2). A peak at 960 ± 5 cm^{-1} is likely phosphate, although perchlorates cannot be excluded (Fig. 2).

Garde detail scans (*8*) exhibit strong fluorescence signatures, centered at ~340 nm, that spatially correlate with carbonate, probable phosphate, and amorphous silicate spectra localized within narrow intergranular spaces (Fig. 2, E and F). A less-intense fluorescence band centered at 285 nm typically accompanies the 340-nm fluorescence. Other areas exhibit no fluorescence (Fig. 2C).

Guillaumes features white and reddish-brown patches, 1 to 2 mm across, that do not have well-formed crystal faces (Fig. 3A and fig. S1). These are secondary materials within a basaltic igneous rock (*9*), which we interpret as void fills, that correlate with sulfate and perchlorate spectra. Spectra with high-intensity 950- to 955-cm^{-1} peaks and minor 1090- to 1095-cm^{-1} and 1150- to 1155-cm^{-1} peaks match laboratory measurements of anhydrous Na-perchlorate (*8*) (Fig. 3G and fig. S4). Two strong Na-perchlorate detections correlate with centers of the brightest material within the patches (*8*). Guillaumes spectra commonly contain a single low-intensity peak positioned at 950 to 955 cm^{-1}. We interpret these as low-intensity Na-perchlorate peaks, although the cation species is uncertain because of a lack of resolvable minor peaks (*8*) (Fig. 3G). Other spectra exhibit both 950- to 955-cm^{-1} peaks and equally strong 1010- to 1020-cm^{-1} peaks, with low-intensity broad features at 1120 ± 5 cm^{-1}, and occasional broad 3450 ± 5 cm^{-1} hydration features, consistent with a mixture of sulfate and perchlorate that is minimally hydrated (Fig. 3G). A Ca-sulfate species best explains these spectra when coupled with elemental chemistry data from PIXL and SuperCam (*8*). Two detected 965 cm^{-1} peaks are likely phosphates, although perchlorates cannot be excluded.

Bellegarde contains white 0.5- to 1-mm secondary crystals with well-formed and sharp crystal faces and reddish-brown semi-isopachous rims (*8*) (fig. S2). We interpret these as void fills within a basaltic igneous rock. These crystals exhibit 1010- to 1020-cm^{-1} peaks, similarly attributed to Ca-sulfate when coupled with elemental chemistry data (*8*). Several of the sulfate peaks are also associated with a narrow low-intensity hydration feature at 3560 ± 5 cm^{-1}, consistent with hydrated Ca-sulfates (fig. S3). The Bellegarde target contains a single 1080 ± 5 cm^{-1} peak of possible Ca-carbonate (Fig. 4). Narrow peaks at 975 cm^{-1} could not definitively be identified and could be phosphate or perchlorate (*8*). The SHERLOC mineral detections within the Bellegarde and Guillaumes targets are consistent with the results from other Perseverance instruments (*8*, *9*, *11*) (figs. S5 and S6).

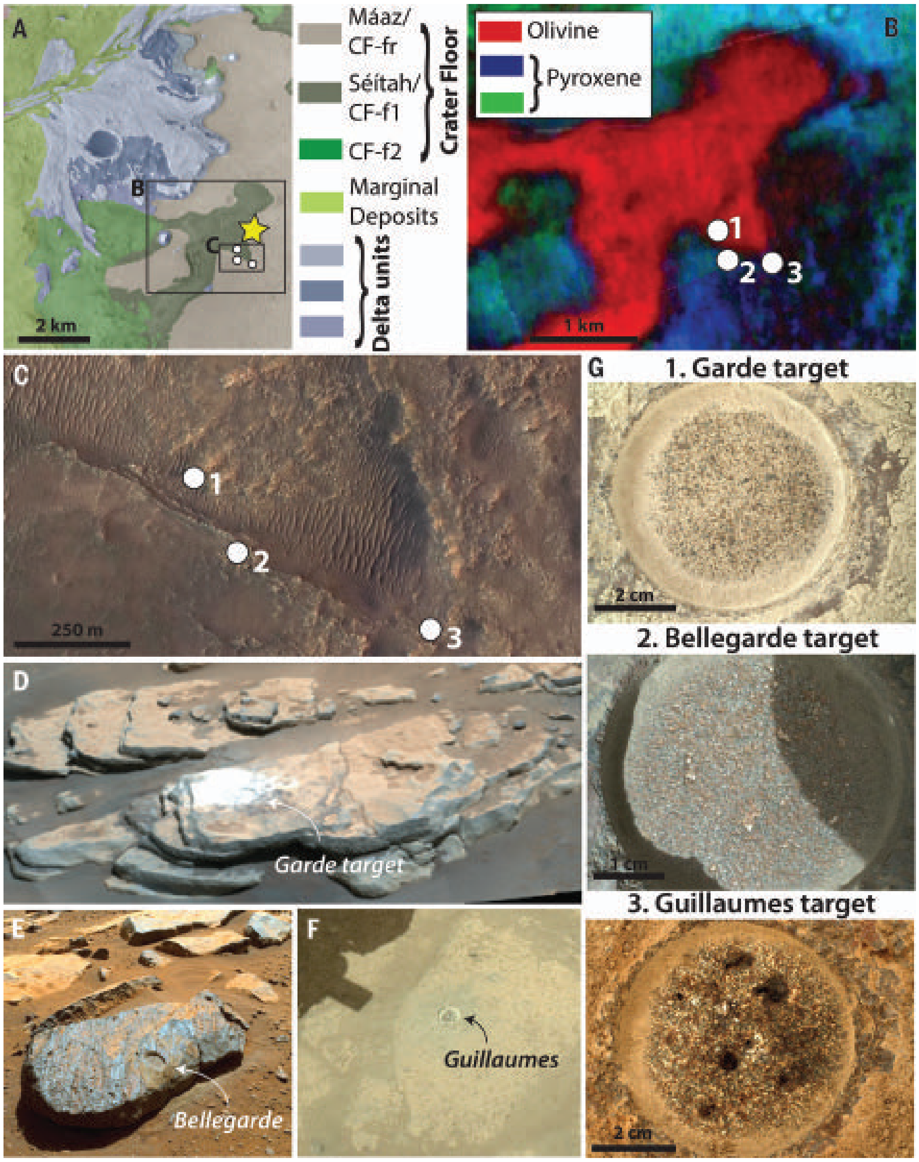


Fig. 1. Rover images of the three abraded targets and their orbital context. (**A**) Map of orbitally defined geological units within Jezero crater (*12*), including the Crater Floor Fractured Rough unit (CF-fr) equivalent to the Máaz Fm and the Crater Floor Fractured 1 (CF-f1) unit equivalent to the Séítah Fm. The star shows the landing site, and white circles show the positions of the three abraded targets. The black outlines indicate the locations shown in (B) and (C). [Adapted by permission from Springer Nature Customer Service Center GmbH, Springer Nature *Space Sci. Rev.* (*12*), copyright (2020)] (**B**) Orbital infrared spectroscopy map showing the location of pyroxene- or olivine-bearing materials in the study area from (*4*). The numbered white circles correspond to the images shown in (G). (**C**) High Resolution Imaging Science Experiment (HiRISE) view of the study area (*30*). The numbered white circles correspond to the images shown in (G). (**D**) Mastcam-Z image showing the Garde patch on the Bastide outcrop. (**E**) Hazard avoidance camera (Hazcam) image showing the Bellegarde patch on the Rochette rock. (**F**) Navigational camera (Navcam) image showing the Guillaumes patch on the Roubion outcrop. (**G**) WATSON images of abraded targets analyzed in this study. Grayscale images are available in figs. S9 to S11.

Guillaumes and Bellegarde targets commonly exhibit a weak, broad fluorescence feature with a maximum at ~340 nm (Figs. 3D and 4D) that appears to be widely distributed across each surface and is occasionally correlated with reddish-brown materials. Although this feature sometimes co-occurs with perchlorate, sulfate, and possibly phosphate, it occurs equally often in areas with unidentified mineralogy. Bellegarde has two other signatures at ~275 and ~305 nm (Fig. 4D), which are strong and localized on specific, light-toned features. The ~305-nm signature is associated with detected sulfate (Fig. 4, D to F, and fig. S2).

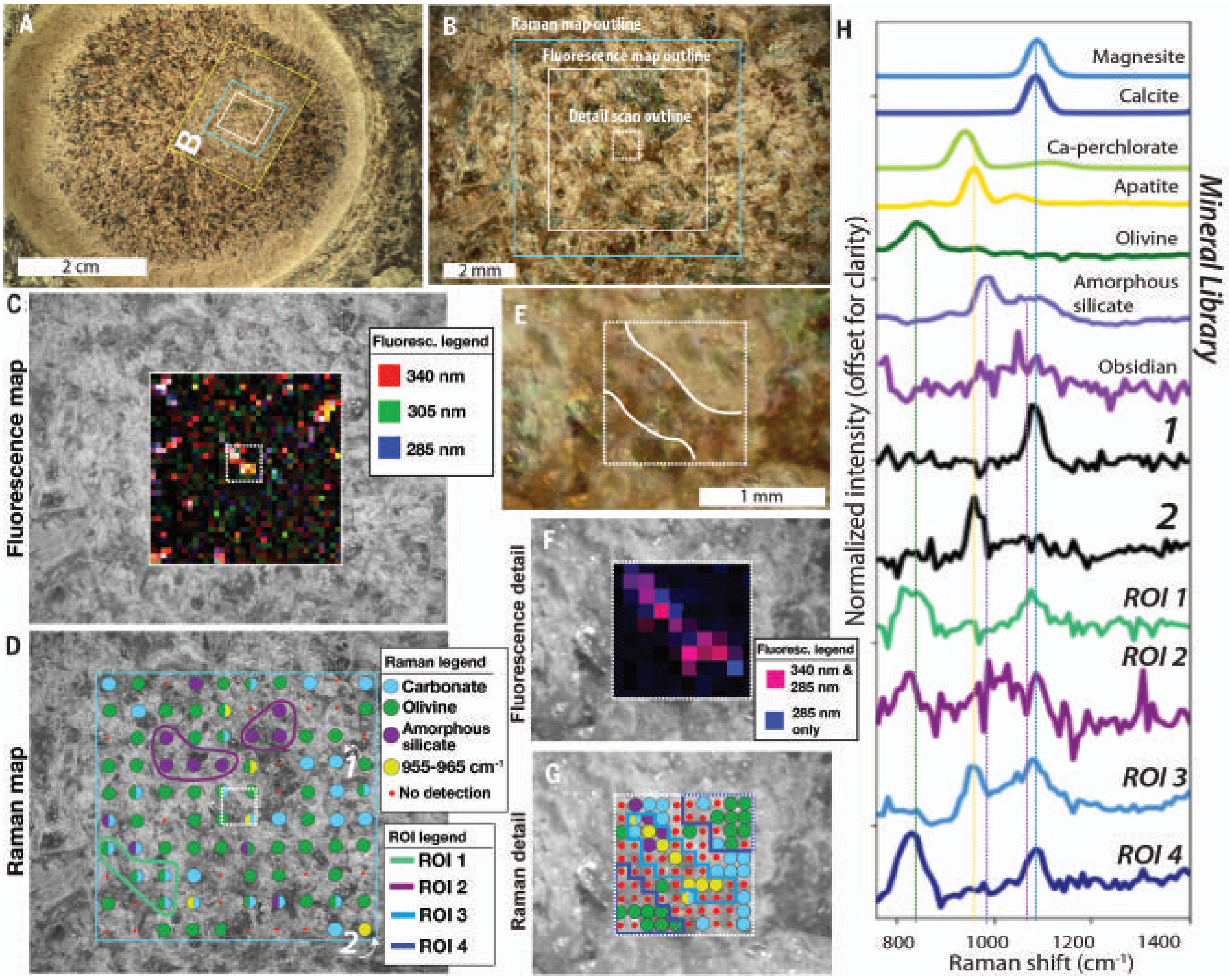


Fig. 2. SHERLOC Raman and fluorescence results for the Garde abraded patch. (**A**) WATSON image. The yellow outline indicates the region shown in (B). (**B**) Context image and scan outlines. Grayscale versions of the context image with data superimposed are shown in (C), (D), (F), and (G). (**C**) Fluorescence map showing the intensity of three main features centralized at 340, 305, and 285 nm in red, green, and blue, respectively. (**D**) Raman mineral maps showing the location of detected olivine, carbonate, phosphate, and weak amorphous silicate features. Purple and green outlines indicate ROIs 1 and 2, numbered in white, which were used for the spectra shown in (H). (**E**) Magnified view of (B) showing fluorescence correlation with intergranular spaces (outlined in white). (**F**) Fluorescence map from the detail scan. (**G**) Raman map from the detail scan and ROIs 3 and 4 outline intergranular and mineral domain textures [same legend as in (D)] used in (H). (**H**) Average ROI and single point (1 and 2) SHERLOC spectra [positions shown in (D) and (G)] compared with laboratory measurements. The vertical dashed lines indicate the peak positions of matching laboratory spectra. Grayscale images are available in figs. S12 to S15.

In Guillaumes, a second fluorescence signature at ~275 nm (Fig. 3D) is observed in two locations, which are ~300 μm in diameter, and are coincident with previous SuperCam laser spots (*8*).

Observation of olivine and carbonate mixtures within the Garde target of the Séítah Fm is consistent with orbital infrared observations (*2*–*4*) and substantiated by multiple lines of evidence (*9*–*11*). Previously proposed hypotheses for the precipitation of these carbonates include low- and high-temperature aqueous alteration of olivine-rich igneous materials, which we will subsequently refer to as ultramafic protolith (*3*, *15*–*17*), or precipitation from lake or groundwater (*4*, *15*–*17*). Our 10- to 100-μm-scale textural and spectroscopic evidence supports carbonate formation through aqueous alteration of an ultramafic protolith, known as carbonation. The supporting evidence includes the following: (i) Carbonate cation compositions are consistent with those of olivines, suggesting that mixed Fe- and Mg-olivine gave rise to mixed Fe- and Mg-carbonates, similar to carbonation environments on Earth and within martian meteorites (*16*, *18*, *19*). (ii) The observed carbonates co-occur with hydrated materials (*9*) and potentially aqueously formed amorphous silicates and phosphate. (iii) The spectral and textural variation of olivine- and carbonate-dominated zones within both primary grains and intergranular spaces are expected for carbonated ultramafic protoliths on Earth (*16*) and within martian meteorites (*18*, *19*).

These observations suggest that the degree of aqueous alteration to the ultramafic protolith was not pervasive, because large olivine-rich domains remain intact. By contrast, the alteration of the primary lithology is pervasive and occurs throughout (not in specific spatial domains, for example, fractures). In ultramafic alteration environments on Earth (*16*) and in martian meteorites (*18*, *19*), carbonation can be associated with the formation of oxides, hydroxides, and/or Fe- and Mg-rich phyllosilicates, which have not been observed (*9*). Carbonation can occur under a wide range of temperatures from low to ambient to hydrothermal and metamorphic (*15*–*17*). Other alteration minerals, such as serpentine, have not been definitively observed in the Séítah Fm to

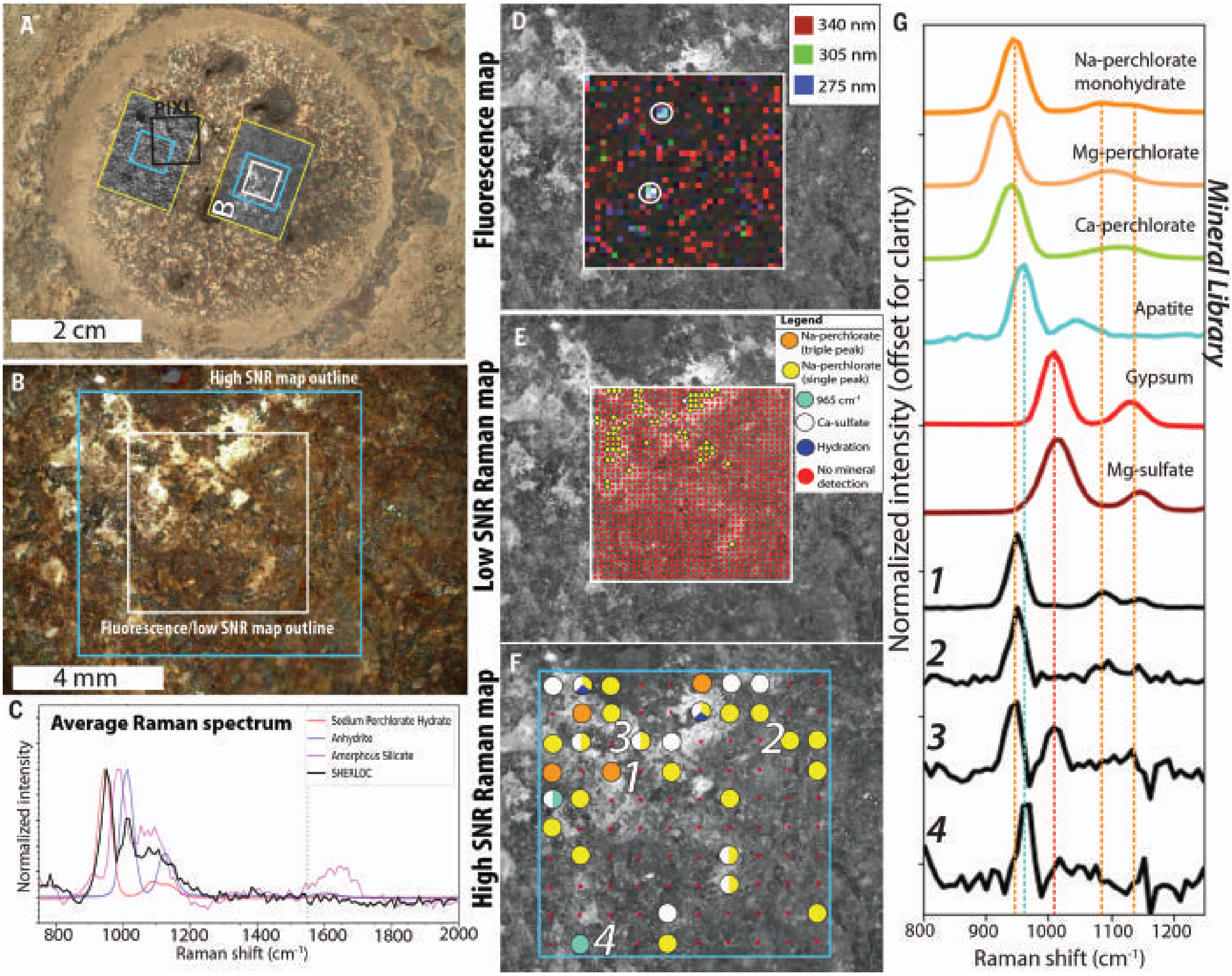


Fig. 3. SHERLOC Raman and fluorescence results for the Guillaumes abraded patch. (**A**) WATSON image. Two SHERLOC scans (yellow outlines) and one PIXL scan (black outline) are shown. The yellow outline on the right indicates the region shown in (B). (**B**) Context image and scan outlines. Grayscale versions of the context image with data superimposed are shown in (D) to (F). SNR, signal-to-noise ratio. (**C**) Average Raman spectrum compared with laboratory measurements of amorphous silicate, Na-perchlorate, and anhydrite. Laboratory spectral features at 1500 to 1600 cm^{-1} are O_2 (vertical, dotted line) and organic contaminants. (**D**) Fluorescence map showing the intensity of three main features centralized at 340, 305, and 275 nm in red, green, and blue, respectively. The white circles indicate the locations exposed to SuperCam laser shots (*8*). (**E** and **F**) Low- and high-SNR (*8*) Raman mineral maps showing the locations of detected perchlorate, Ca-sulfate with and without hydration, and 965 cm^{-1} peaks. The numbers correspond to the spectra shown in (G). (**G**) Raman spectra from SHERLOC (1 to 4) [positions indicated with numbers in (E) and (F)] compared with laboratory measurements. The vertical dashed lines indicate the peak positions of matching laboratory spectra. Grayscale images are available in figs. S16 to S19.

date, which could suggest time-limited interactions, low water-to-rock ratios, or ambient fluid temperatures during carbonation (*3*, *15*–*17*).

The similarity between the mineralogy of the Garde target in the Séítah Fm to the surrounding widespread, regional olivine-carbonate-bearing unit with a similar orbital spectroscopic signature and geomorphological texture (*3*, *4*, *15*, *16*) suggests that carbonation of olivine may have occurred throughout this extensive region on ancient Mars (~2.7 billion to 3.8 billion years ago). These observations parallel those made by the Spirit rover in Gusev crater (*20*) and within (1.3 billion to 4 billion years ago) martian meteorites (*18*, *19*). Modeling has suggested that carbonate deposition could have played a role in the evolution of Mars' atmosphere (*3*, *17*, *21*), but the geological nature of such a depositional mechanism had remained unexplained. Taken together, micron-scale SHERLOC observations of this phenomenon complement previous orbital and meteorite observations and demonstrate the alteration of igneous materials that results in geological deposition of carbonates.

Jezero crater perchlorate detections, similar to those found by the Phoenix lander (*22*), have been substantiated using three independent instruments (*9*). Previous evidence for martian perchlorates also includes observations by the Curiosity rover (*23*), proposed but later disputed orbital detections (*24*), and detection within the Tissint meteorite (*25*). The SHERLOC perchlorate detections differ from previous mission observations because they are observed within the interior of a rock and not on the surface, are related to aqueous processes, and are likely Na-perchlorate [not previously detected Ca-, Fe-, or Mg-perchlorates (*23*)].

Previous hypotheses for perchlorate formation on Mars are (i) irradiation of chlorine-bearing parent minerals (*26*), (ii) atmospheric

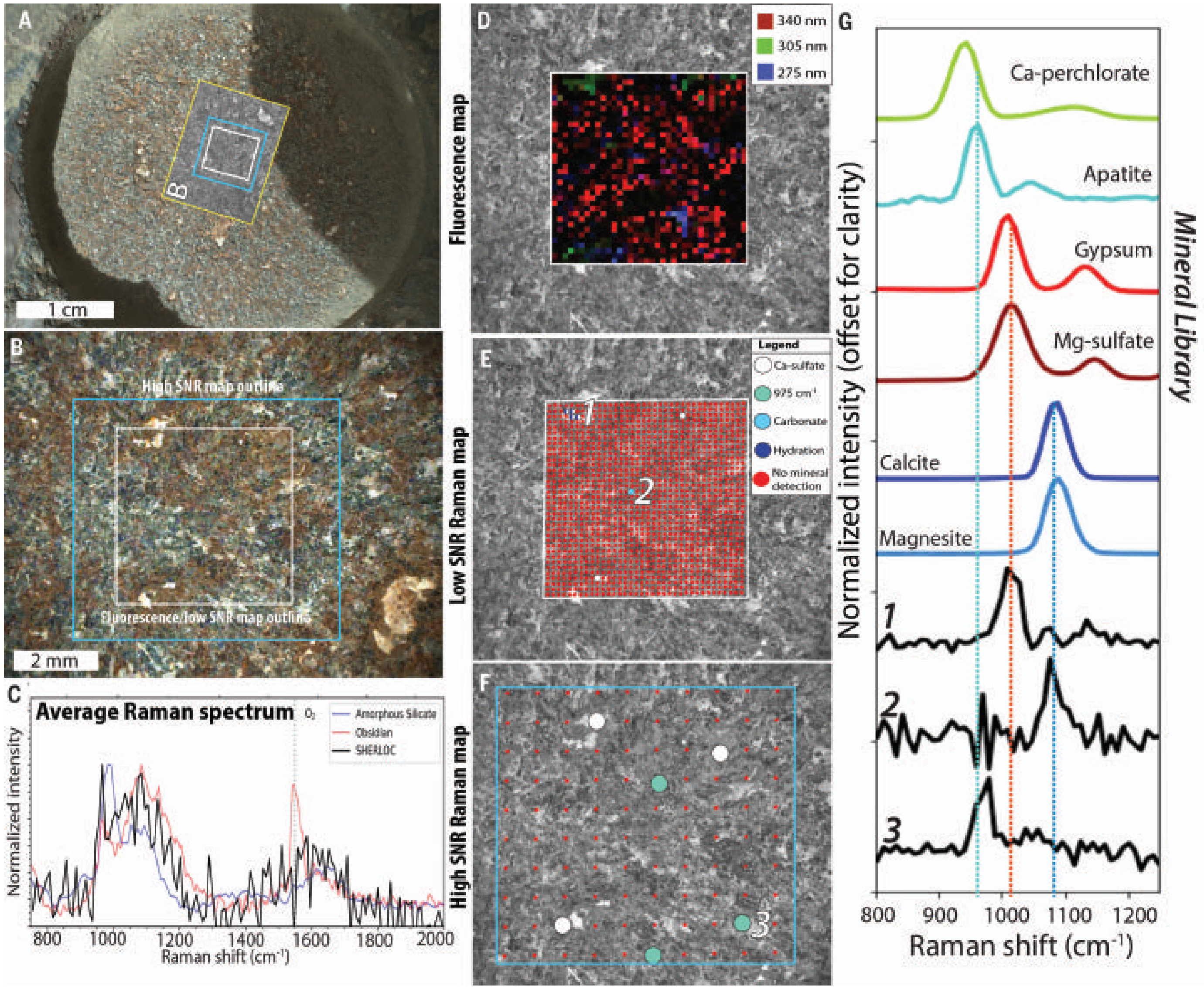


Fig. 4. SHERLOC Raman and fluorescence results for the Bellegarde abraded patch. (**A**) WATSON image. The yellow outline indicates the region shown in (B). (**B**) Context image and scan outlines. Grayscale versions of the context image with data superimposed are shown in (D) to (F). (**C**) Average Raman spectrum compared with laboratory measurements of amorphous silicate and obsidian. Laboratory spectral features at 1500 to 1600 cm^{-1} are O_2 (vertical, dotted line) and trace organic contaminants. (**D**) Fluorescence map showing the intensity of three main features centralized at 340, 305, and 275 nm in red, green, and blue, respectively. (**E** and **F**) Low- and high-SNR (*8*) Raman mineral maps showing the location of detected Ca-sulfate with and without hydration, carbonate, and 975 cm^{-1} peaks. The numbers correspond to the spectra shown in (G). (**G**) Raman spectra from SHERLOC (1 to 3) [positions indicated with numbers in (E) and (F)] compared with laboratory measurements [full hydrated sulfate spectrum in fig. S3 (*8*)]. The vertical dashed lines indicate the peak positions of matching laboratory spectra. Grayscale images are available in figs. S20 to S23.

oxidation of chlorine species (*27*), or (iii) formation from brines (*25*). Perchlorates could also be mobilized in thin films of fluid (*23*). The Jezero perchlorates occur in white void fills within rock interiors and did not form directly on the surface as expected from materials formed by cosmic irradiation or atmospheric oxidation, indicating either formation or mobilization through briny fluids after basalt formation. The concomitant detection of sulfates and perchlorates within the Guillaumes target suggests that sulfate formed together with perchlorate or parent chlorine-bearing species, such as halite, within percolating briny waters that were then subsequently oxidized to perchlorate. Bellegarde exhibited sulfate without perchlorate, suggesting that these brines did not precipitate chlorine-bearing parent species, that perchlorate formation was not pervasive, or that perchlorates have since been dissolved. Perchlorates are easily dissolved, and therefore perchlorates likely formed when these rocks were last exposed to liquid water. Perchlorate and sulfate detections within the stratigraphically younger Máaz Fm (*9*) substantiates an aqueous environment on Mars that occurred separately from the stratigraphically older (*9*) Séítah Fm carbonation environment.

Deep-ultraviolet fluorescence is particularly sensitive to aromatic organic compounds, and the fluorescence signatures observed in all three targets are consistent with emission from aromatic organic compounds that contain one or two fused aromatic rings and/or aromatic heterocycles (*7*, *8*, *28*) (fig. S7). Although fluorescence signatures cannot be assigned to

specific organic compounds, the ~340-nm fluorescence is consistent with a base structure of two-ring aromatic organics like naphthalene, whereas ~275- to 285-nm fluorescence is more consistent with one-ring aromatic organics such as benzene (*8*, *28*). The ~305-nm fluorescence could indicate either one- or two-ring aromatics, depending on functional groups. We interpret ~305- and ~275-nm fluorescence as organics that occurred with sulfates within the Bellegarde target (Fig. 4, D to F), whereas we interpret ~285-nm fluorescence as organics that occurred with carbonate-phosphate-amorphous silicate brown microcrystalline alteration zones within the Garde target (Fig. 2) (*8*). The ~340-nm fluorescence co-occurs with carbonate-phosphate-amorphous silicate alteration zones in Garde but is not associated with particular phases in Guillaumes and Bellegarde (*8*). This phosphate-correlated 340-nm fluorescence could be explained by organics in the same alteration zone and/or a minor component of fluorescent cerium present within the phosphate (*8*). In Guillaumes and Bellegarde, 340-nm fluorescence is predominantly uncoupled from phosphate detections, suggesting that part of this signal is best assigned to organics (*8*). However, that some or all of the ~340-nm fluorescence signal is from cerium cannot be excluded (*8*).

When combined with Curiosity observations of organic material in mudstones (*29*), the presence of organic material in igneous rocks would imply a diverse relationship between geological processes and organic compounds on Mars. Several explanations for the presence of martian organics are possible, for example, infall of meteoritic material (*6*), in situ synthesis mechanisms (*18*, *19*, *25*), or a putative relic martian biosphere. The association between organics and sulfate-, phosphate-, perchlorate-, carbonate-, and amorphous silicate–bearing mineralogy, as well as alteration textures, suggests that aqueous alteration of igneous rocks could have played a role in the preservation (or synthesis) of these organics, as seen for similar organics-mineral correlations in martian meteorites (*18*, *19*, *25*). However, potentially organic, widespread ~340-nm fluorescence could suggest that other processes, such as dust, contributed to detections. Some mineral phase associations might not be apparent because of instrumental limitations (*8*).

We did not detect Raman peaks consistent with aromatic organic compounds, such as the C=C stretching mode (or G band) around ~1600 cm^{-1}. However, Raman scattered light is several orders of magnitude weaker than fluorescence (*8*, *28*). Organic concentrations were possibly insufficient to produce detectable Raman scattering, either because of low original concentration or subsequent degradation. We estimate a range from 5×10^{-11} to 3×10^{-10} g of aromatic organics in localized scan points that encompass an estimated 6×10^{-7} g of rock (*8*). Estimates from the average fluorescence maps suggest a bulk concentration of 0.1 to 10 parts per million (ppm), with higher concentrations associated with more aqueously altered surfaces (*8*). This is consistent with known bulk concentrations of organics that contain one- and two-ring aromatic species and are indigenous to martian meteorites [11.2 ± 6.9 ppm (*18*)] and Curiosity rover detections in mudstones [~70 parts per billion by weight to 10.6 ± 8.9 ppm (*6*, *29*)].

Collectively, these data show that the drilled samples collected by Perseverance from the floor of Jezero crater are likely to contain evidence for carbonation and formation of sulfates and perchlorates. Fluorescence signatures consistent with organics that are present within these materials indicate an interplay between igneous rocks, aqueous alteration, and organic material on Mars.

REFERENCES AND NOTES

1. K. A. Farley *et al.*, *Space Sci. Rev.* **216**, 142 (2020).
2. T. A. Goudge, J. F. Mustard, J. W. Head, C. I. Fassett, S. M. Wiseman, *J. Geophys. Res. Planets* **120**, 775–808 (2015).
3. B. L. Ehlmann *et al.*, *Science* **322**, 1828–1832 (2008).
4. B. H. Horgan, R. B. Anderson, G. Dromart, E. S. Amador, M. S. Rice, *Icarus* **339**, 113526 (2020).
5. R. Navarro-González, E. Vargas, J. de la Rosa, A. C. Raga, C. P. McKay, *J. Geophys. Res. Planets* **115**, E12010 (2010).
6. J. L. Eigenbrode *et al.*, *Science* **360**, 1096–1101 (2018).
7. R. Bhartia *et al.*, *Space Sci. Rev.* **217**, 58 (2021).
8. Materials and methods are available as supplementary materials.
9. K. A. Farley *et al.*, *Science* **377**, eabo2196 (2022).
10. Y. Liu *et al.*, *Science* **377**, 1513–1519 (2022).
11. M. M. Tice *et al.*, *Sci. Adv.* **8**, eabp9084 (2022).
12. K. M. Stack *et al.*, *Space Sci. Rev.* **216**, 127 (2020).
13. J. Razzell Hollis *et al.*, *Planet. Space Sci.* **209**, 105356 (2021).
14. J. Razzell Hollis, S. Ireland, W. Abbey, R. Bhartia, L. W. Beegle, *Icarus* **357**, 114067 (2021).
15. J. D. Tarnas *et al.*, *J. Geophys. Res. Planets* **126**, JE006898 (2021).
16. E. L. Scheller *et al.*, *J. Geophys. Res. Planets* **126**, JE006828 (2021).
17. V. F. Chevrier, M. Morisson, *J. Geophys. Res. Planets* **126**, e2020JE006698 (2021).
18. A. Steele, F. M. McCubbin, M. D. Fries, *Meteorit. Planet. Sci.* **51**, 2203–2225 (2016).
19. A. Steele *et al.*, *Science* **375**, 172–177 (2022).
20. R. V. Morris *et al.*, *Science* **329**, 421–424 (2010).
21. R. Hu, D. M. Kass, B. L. Ehlmann, Y. L. Yung, *Nat. Commun.* **6**, 10003 (2015).
22. M. H. Hecht *et al.*, *Science* **325**, 64–67 (2009).
23. P. E. Martin *et al.*, *J. Geophys. Res. Planets* **125**, e2019JE006156 (2020).
24. E. K. Leask, B. L. Ehlmann, M. M. Dundar, S. L. Murchie, F. P. Seelos, *Geophys. Res. Lett.* **45**, 12180–12189 (2018).
25. A. Steele *et al.*, *Sci. Adv.* **4**, eaat5118 (2018).
26. B. L. Carrier, S. P. Kounaves, *Geophys. Res. Lett.* **42**, 3739–3745 (2015).
27. D. C. Catling *et al.*, *J. Geophys. Res. Planets* **115**, E1 (2010).
28. R. Bhartia *et al.*, *Appl. Spectrosc.* **62**, 1070–1077 (2008).
29. C. Freissinet *et al.*, *J. Geophys. Res. Planets* **120**, 495–514 (2015).
30. USGS Astrogeology Science Center, Mars 2020 terrain relative navigation HiRISE orthorectified image mosaic, *Astropedia* (2020); https://doi.org/10.5066/P9QJDP48.
31. L. W. Beegle, R. Bhartia, Mars 2020 SHERLOC Bundle, *Planetary Data System* (2021); https://doi.org/10.17189/1522643.
32. J. F. Bell, J. N. Maki, Mars 2020 Mast Camera Zoom Data Bundle, from Operations Team, calibrated products, *Planetary Data System* (2021); https://doi.org/10.17189/bs6b-4782.
33. J. N. Maki, Mars 2020 Engineering Camera (ECAM) Bundles, *Planetary Data System* (2021); https://doi.org/10.17189/1522847.
34. A. C. Allwood, J. A. Hurowitz, Mars 2020 PIXL Raw and Processed Data Bundle, *Planetary Data System* (2021); https://doi.org/10.17189/1522645.
35. R. C. Wiens, M. A. Sylvestre, Mars 2020 SuperCam Bundle, *Planetary Data System* (2021); https://doi.org/10.17189/1522646.
36. K. Uckert, LoupeV5.1.5, *Zenodo* (2022); https://doi.org/10.5281/zenodo.7062998.
37. E. L. Scheller, Laboratory datasets - Raman spectra, *Zenodo* (2022); https://doi.org/10.5281/zenodo.7212486.

ACKNOWLEDGMENTS

We thank the entire Perseverance rover team. The work described in this paper was partially carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. We thank K. Edgett for his contributions to the team and S. Le Mouélic for helpful comments. We thank V. Chevrier and two anonymous reviewers for their improvements to the manuscript. **Funding:** E.L.S. was supported by a NASA Earth and Space Science Fellowship (NESSF) (grant 80NSSC18K1255). J.R.H. and A.C.F. were supported by a NASA Postdoctoral Program fellowship. E.L.S., J.R.H., A.S., L.W.B., R.B., B.L.E., P.C., M.F., F.M.M., and A.S.B. were supported by the NASA Mars 2020 Phase-E funds to the SHERLOC investigation. A.J.W. was supported by the NASA M2020 Participating Scientist Program. T.F. was supported by an Italian Space Agency grant (no. 2017-48-H.0). A.G.F. was supported by a European Research Council Consolidator Grant (no. 818602). R.C.W. was funded by NASA contracts NNH15AZ25I and NNH13ZDA018O. S.Si. was supported by the Swedish National Space Agency (contracts 137/19 and 2021-00092). K.H.-L. is funded by an Aurora Research Fellowship from the UK Space Agency (grant ST/V00560X/1). **Author contributions:** E.L.S. and J.R.H. contributed equally to data analysis. E.L.S. and J.R.H. wrote the manuscript with substantial contributions from A.S. and E.C. L.W.B. and R.B. are the principal and deputy investigators of the SHERLOC instrument. K.U. developed the Loupe software. P.C., S.Sha., B.L.E., W.J.A., S.A.A., E.L.B., B.L.B., A.S.B., S.V.B., L.D., D.M.F., T.F., A.C.F., M.F., K.H.-L., W.F.H., J.E.H., S.I., R.S.J., M.R.K., C.L., F.M.M., B.E.N., C.R.S.-V., R.D.R., S.Si., S.Shk., K.S., A.W., K.H.W., K.W., B.W., A.Y., and R.Y. are members of the SHERLOC science and operations teams, who participated in planning and carrying out SHERLOC data acquisitions, calibration, and processing during rover operations. The SHERLOC science team also provided laboratory measurements and data analysis and contributed to the writing of the initial manuscript. T.K., K.R.M., and Y.L. are members of the PIXL team, who assisted in paired SHERLOC and PIXL data analysis. K.C.B., O.B., T.B., A.J.B., E.L.C., A.G.F., K.A.F., L.C.K., P.K., L.M., J.I.N., M.S., M.A.S., S.K.S., D.L.S., J.I.S., R.J.S., K.M.S., B.P.W., A.J.W., R.C.W., and M.P.-Z. are members of the wider Perseverance rover science team; they contributed to planning the SHERLOC data acquisitions, participated in data interpretation, and/or assisted in revising the manuscript. **Competing interests:** We declare no competing interests. **Data and materials availability:** The data are available on the Planetary Data System, in the bundles for SHERLOC and WATSON (*31*), Mastcam-Z (*32*), Navcam and Hazcam (*33*), PIXL (*34*), and Supercam (*35*); the exact files we used are listed in the supplementary materials. The Loupe software is archived at Zenodo (*36*). Laboratory data used for the study are archived at Zenodo (*37*).

License information: Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abo5204
Materials and Methods
Figs. S1 to S23
Table S1
References (*38*–*65*)

Submitted 15 February 2022; accepted 10 November 2022
Published online 23 November 2022
10.1126/science.abo5204

ANTIBIOTIC RESISTANCE

Tuberculosis treatment failure associated with evolution of antibiotic resilience

Qingyun Liu[1]†, Junhao Zhu[1]†, Charles L. Dulberger[1,2]‡, Sydney Stanley[1], Sean Wilson[2]§, Eun Seon Chung[3,4], Xin Wang[1], Peter Culviner[1], Yue J. Liu[1], Nathan D. Hicks[1]¶, Gregory H. Babunovic[1], Samantha R. Giffen[1], Bree B. Aldridge[3,4], Ethan C. Garner[2], Eric J. Rubin[1], Michael C. Chao[1], Sarah M. Fortune[1,5]*

The widespread use of antibiotics has placed bacterial pathogens under intense pressure to evolve new survival mechanisms. Genomic analysis of 51,229 *Mycobacterium tuberculosis* (*Mtb*) clinical isolates has identified an essential transcriptional regulator, *Rv1830*, herein called *resR* for resilience regulator, as a frequent target of positive (adaptive) selection. *resR* mutants do not show canonical drug resistance or drug tolerance but instead shorten the post-antibiotic effect, meaning that they enable *Mtb* to resume growth after drug exposure substantially faster than wild-type strains. We refer to this phenotype as antibiotic resilience. ResR acts in a regulatory cascade with other transcription factors controlling cell growth and division, which are also under positive selection in clinical isolates of *Mtb*. Mutations of these genes are associated with treatment failure and the acquisition of canonical drug resistance.

Antibiotics are designed to kill bacteria by inhibiting biological processes essential for their survival. Completely eliminating a population of bacteria with antibiotics is challenging even when the bacteria are nominally sensitive to the drug (*1*). There is a growing body of work defining the mechanisms beyond canonical resistance by which bacterial populations avoid clearance by antibiotics (*2*, *3*). These include the formation of privileged subpopulations of drug-tolerant bacteria, as well as the acquisition of other traits that enhance bacterial survival, including variations in the cellular requirements for the target and alterations in drug penetration or efflux (*2*–*4*). We hypothesized that bacterial adaptation to antibiotics would be recorded as mutations in the genes associated with the most clinically relevant of these adaptive processes, and determining the targets of natural selection would reveal new mechanisms that enable bacteria to survive antibiotics in patients. Therefore, we sought to identify the evolutionary signatures of adaptation to antibiotics in *Mycobacterium tuberculosis* (*Mtb*). This obligate human pathogen has remained one of the largest causes of mortality despite facing months of antibiotic pressure in every treated patient and having been under antibiotic selection globally since the introduction of streptomycin in 1944 (*5*).

[1]Department of Immunology and Infectious Diseases, Harvard T. H. Chan School of Public Health, Boston, MA 02115, USA. [2]Department of Molecular and Cellular Biology, Harvard University, Boston, MA, USA. [3]Department of Molecular Biology and Microbiology, Tufts University School of Medicine, Boston, MA 02111, USA. [4]Department of Biomedical Engineering, Tufts University School of Engineering, Medford, MA 02115, USA. [5]Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA.
*Corresponding author. Email: sfortune@hsph.harvard.edu
†These authors contributed equally to this work.
§Present address: California Institute of Technology, Pasadena, CA, USA.
¶Present address: Seres Therapeutics, Cambridge, MA, USA.

Ongoing positive selection in the *Mtb* population

When clinical *Mtb* isolates are sequenced, the specimen is typically not derived from a single colony, but rather represents the bacterial population sampled from the patient. This population contains mutations arising from within-host evolution of the bacterium, which can be tracked by identifying unfixed mutations detected in deep sequencing (Fig. 1A) (*6*, *7*). We reasoned that by looking at within-host evolution across very large numbers of clinical isolates, even in the absence of clinical metadata, we could identify genes that repeatedly acquire mutations (parallel evolution) within different patients that would mark processes under positive selection. Therefore, we assembled previously published whole-genome sequencing data from 51,229 *Mtb* clinical isolates (Fig. 1B) and used a mutation burden test and the ratio of nonsynonymous to synonymous mutations (*dN/dS*) to identify genes that are under positive selection in the *Mtb* population (see the materials and methods). We also identified intergenic regions (IGRs) that are highly enriched for mutations. As expected, genes and IGRs previously associated with drug resistance or tolerance were highly mutated and the genes had *dN/dS* ratios >1, which is indicative of positive selection (Fig. 1B and fig. S1). We also identified a second set of genes that showed similar selective signals but for which the selective pressures are unknown (Fig. 1B).

Functional enrichment analysis indicated that the genes under selection are highly enriched for transcriptional regulators (12/60, $P = 1.57 \times 10^{-5}$; Fig. 1, B and C). One of the most frequently mutated genes is *resR* (gene identifier: *Rv1830*), an essential gene that is predicted to be a *merR*-type regulator. We also found evidence of positive selection in an analysis of the fixed mutations in *resR* (*dN/dS* = 2.91). The pattern of mutation was very similar between unfixed and fixed mutations (fig. S2), suggesting that within-host selection leads to mutation fixation. The *dN/dS* of *resR* was higher in drug-resistant strains compared with drug-sensitive strains (Fig. 1D), suggesting that *resR* mutations are associated with the evolution of drug resistance and consistent with a recent study showing that *resR* mutations were associated with a multidrug-resistant tuberculosis (TB) phenotype (*8*). Protein structural modeling indicated that most *resR* clinical mutations occur in the *merR*-type helix-turn-helix DNA-binding domain, with the most common mutations located at the interface between ResR and duplex DNA (Fig. 1E and fig. S2), where they could affect DNA-binding affinity and the transcription of downstream genes.

resR mutants display antibiotic resilience

To investigate the functional consequence of *resR* mutations, we used oligo-mediated recombineering to make point mutant strains by individually introducing three common clinical mutations (P59L, A85V, and R95C) into the chromosomal copy of *resR* in *Mtb* (H37Rv) (Fig. 1F). These *resR* mutations did not affect bacterial growth under standard or host-relevant carbon source conditions (fig. S3, A and B). Given the association with drug resistance, which occurs through de novo chromosomal mutation in *Mtb*, we also assessed the effect of *resR* mutations on the bacterial mutation rate, but did not find alterations (fig. S3C and table S1). However, we noted that *resR* mutants were ~20% longer and ~5% wider than the wild-type cells (fig. S4) and had a thickened cell envelope (fig. S5), suggesting that mutations of *resR* have functional consequences on bacterial size control under standard growth conditions.

We next tested the effects of *resR* mutations on drug resistance by measuring the minimum inhibitory concentration (MIC) for a panel of eight antibiotics, including first-line, second-line, and new anti-TB drugs (Fig. 2A). *resR* mutants and wild-type strains had very similar MIC profiles, with only subtle MIC shifts identified: a small increase for isoniazid (0.01 to 0.02 μg/ml) and minor decreases for rifampicin and ofloxacin (Fig. 2A and fig. S6). These changes are far below the MIC designated for being classified as drug resistant (Fig. 2A) (*9*). We next assessed the effects of *resR* mutations on drug tolerance by performing time-kill analyses for the eight antibiotics using drug concentrations 100-fold MIC. The *resR* mutants and wild-type strains showed very similar time-kill dynamics, although again, there were subtle differences for some drugs (ofloxacin,

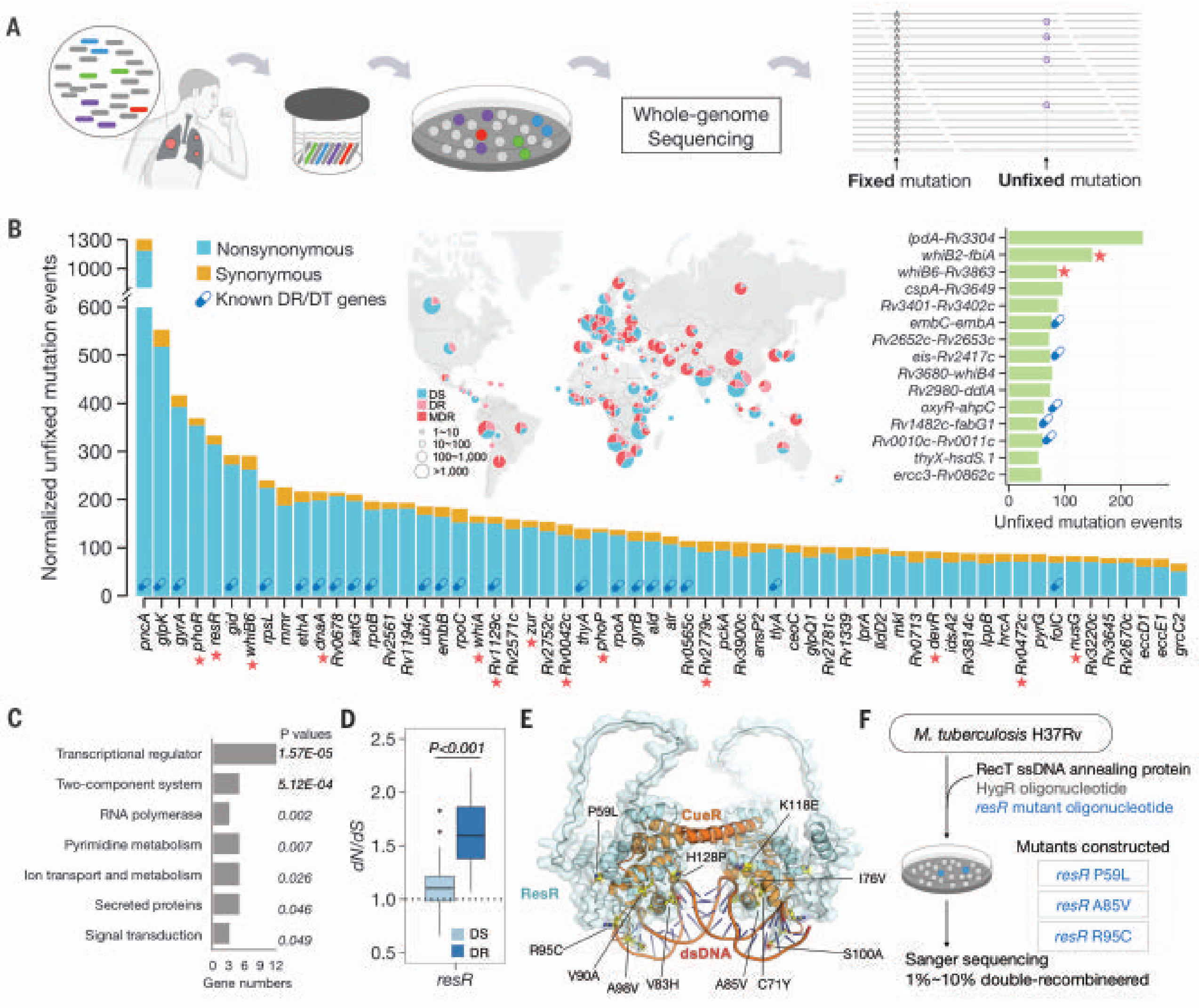


Fig. 1. Ongoing positive selection in the global *Mtb* population. (**A**) Diagram of fixed and unfixed variants. (**B**) Genes and IGRs with signal of ongoing positive selection. Known drug-resistance (DR) or drug-tolerant (DT) genes are marked with capsule symbols. Red stars indicate transcriptional regulators. The map shows the geographic origin of the clinical *Mtb* isolates and their DR status. MDR, multidrug resistance. (**C**) Categories of genes under positive selection with adjusted *P* values for enrichment. (**D**) *dN/dS* ratio of *resR* in DR and DS strains. *P* value was determined by unpaired *t* test. (**E**) Structural model of the ResR dimer (cyan, predicted by Alphafold) aligned to the MerR-family homolog CueR (orange) in complex with duplex DNA with common ResR clinical mutations indicated (yellow). (**F**) Schematic of construction of mutations in the chromosomal copy of *resR* in *Mtb*.

ethionamide, and bedaquiline) at particular time points (Fig. 2B). The changes in the minimum duration of time to kill 99% of bacteria (MDK_{99}) were <24 hours (Fig. 2B); previously characterized drug-tolerance mutants in *Mtb* prolonged the MDK_{99} by more than 4 days (*10*).

In contrast to these modest changes in MIC and MDK_{99}, in the time-kill assays, we observed a strong and unexpected drug phenotype that clearly distinguished *resR* mutants from the wild-type strains. Once the cells were plated on drug-free media after antibiotic exposure, *resR* mutants formed visible colonies significantly earlier than the wild type (Fig. 2C). This early recovery phenotype of *resR* mutants was observed for all eight antibiotics tested, whereas no difference in growth was seen in the absence of drug exposure (Fig. 2C and fig. S7). Using quantitative image analysis to track the growth of individual colonies over time (Fig. 3A and fig. S8) (*11*), we found that antibiotic treatment caused a 4.7- to 6.8-day delay in colony formation for wild-type *Mtb*; these delays were highly reproducible and characteristic for each drug (Fig. 3B). The *resR* mutants reduced the time to colony formation by 20 to 50% for all eight drugs tested (Fig. 3, B and C). The *resR* mutants also showed faster recovery when challenged by combinations of the first-line anti-TB drugs (fig. S9, A to C). We calculated the growth dynamics of traced colonies over time and found that the expansion rate of visible colonies did not differ between WT and mutant cells (fig. S9D), suggesting that faster recovery occurred at early stages before colony appearance.

A delay in bacterial growth after a period of antibiotic exposure is classically referred to as the post-antibiotic effect (PAE) (*12*). The PAE has been recognized for decades across many bacterial pathogens and is an important factor in the design of treatment regimens (*12*, *13*). To date, the PAE remains mechanistically enigmatic, with recent work favoring a detoxification model in which the cells must export residual drug before growth resumes (*14*). Our results suggest that *Mtb* has mechanisms for post-antibiotic recovery that are under genetic control, and that *resR* mutants optimize this system, accelerating bacterial recovery after a wide variety of antibiotic insults. We

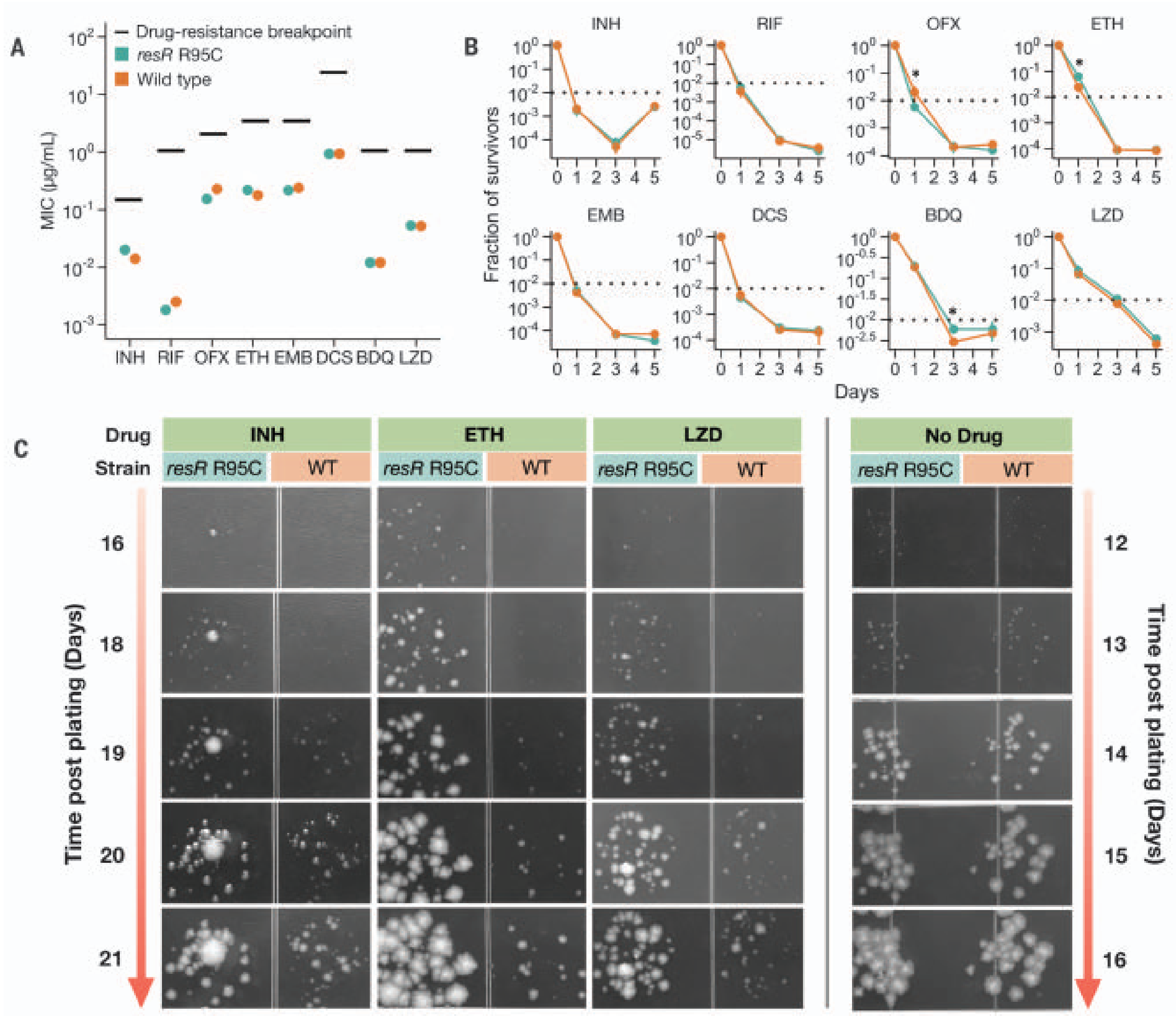


Fig. 2. *resR* mutants show faster recovery after drug exposure. (**A**) MICs of *resR* mutants and wild type for eight anti-TB drugs: INH, isoniazid; RIF, rifampicin; OFX, ofloxacin; ETH, ethionamide; EMB, ethambutol; DCS, D-cycloserine; BDQ, bedaquiline; LZD, linezolid. The drug-resistance breakpoint is the drug concentration that defines clinical drug resistance. (**B**) Time-kill kinetics for eight different drugs for *resR* mutant and wild-type *Mtb* strains. The concentration of each drug was 100-fold MIC. The dashed line indicates MDK_{99}. *$P < 0.05$ by unpaired *t* test. (**C**) Representative images illustrating the post-antibiotic recovery dynamics of *resR* mutant (R95C) and wild-type (WT) *Mtb* strains.

refer to this phenotype as antibiotic resilience and thus named this gene *Rv1830*, resR for resilience regulator.

Antibiotics resilience manifests as early recovery

To better understand early events in post-antibiotic recovery, we used microscopy to investigate cell regrowth after the period of antibiotic exposure. *Mtb* grow from their subpolar regions, which can be quantitatively defined by imaging that tracks the incorporation of fluorescently labeled D-amino acids (FDAAs) into the nascent peptidoglycan near the cell poles (*15*, *16*). We assessed regrowth of wild-type and *resR* mutant cells after 24 hours of isoniazid exposure at 100-fold MIC, followed by a 10- to 36-hour recovery window in drug-free media (Fig. 4, A and B). During the recovery period, *resR* mutants showed significantly more new cell wall synthesis than the wild-type cells ($P < 0.0001$ for 24 and 36 hours; Fig. 4, B and C). Wild-type cells and *resR* mutants had similar FDAA incorporation in the absence of antibiotic exposure (Fig. 4B).

This labeling approach allowed us to use flow cytometry to assess new cell wall synthesis rapidly and quantitatively in mutant and wild-type cells across a wide range of antibiotic exposure conditions. We assessed regrowth after exposure to eight antibiotics at 100-fold MIC. *resR* mutants showed increased cell wall synthesis after exposure to all antibiotics tested, whereas label incorporation was similar between *resR* mutants and the wild type in the absence of drug (Fig. 4D). The different *resR* mutants all showed a similar phenotype (fig. S9E). *resR* mutants also demonstrated faster recovery after up to 120 hours of isoniazid exposure and across a range of antibiotic concentrations (fig. S10).

A regulatory pathway underlying antibiotic resilience

To delineate the regulatory targets through which ResR mediates antibiotic resilience, we used an in vitro whole-genome DNA binding and deep sequencing assay (IDAP-Seq) (*17*) to identify the ResR-binding sites across the genome. We purified ResR proteins with either an N-terminal or a C-terminal His tag, and

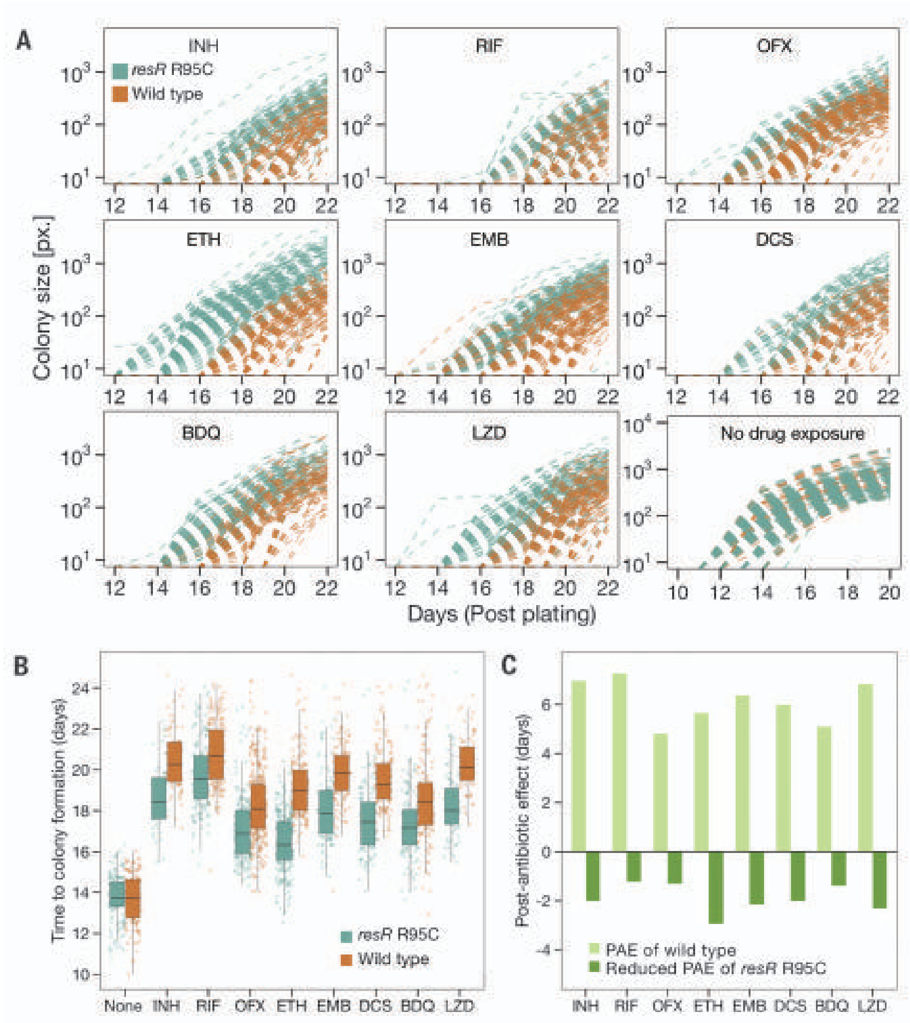


Fig. 3. Quantitative colony size tracking indicates shortened PAE in *resR* mutants. (**A**) Quantitative colony size tracking in pixels (px.) for *resR* mutant and wild-type *Mtb* strains in the presence or absence of antibiotic exposure for 24 hours (24h). (**B**) Duration from plating to the appearance of visible colonies (time to colony formation) for *resR* mutant and wild-type strains after exposure to the indicated antibiotics (24h) or no drug exposure (none). $P = 0.7291$ for the no-drug-exposure group and $P < 0.0001$ for all drug groups, as determined by Mann-Whitney *U* test. (**C**) Bar plot depicting the median values of post-antibiotic delay of wild-type (light green) and *resR* mutant strains (dark green).

performed IDAP-Seq with both versions. With both constructs, we identified four major ResR-binding sites in the *Mtb* genome in the intergenic regions between *rimJ-Rv0996*, *nrdH-Rv3054*, *whiB2-fbiA*, and *Rv3916c-parB* (Fig. 5A). None of the genes flanking the dominant ResR binding sites has a known role in drug resistance. Unexpectedly, however, our population genomic analyses found that one of them, the *whiB2-fbiA* intergenic region, is also one of the major targets of ongoing positive selection in clinical *Mtb* isolates (Fig. 1B). Indeed, the ResR-binding site in the *whiB2-fbiA* intergenic region is the site of mutant accumulation in clinical *Mtb* strains (Fig. 5B), an evolutionary pattern suggesting that natural selection has acted on both the ResR and its binding site, presumably to alter the expression of *whiB2* or *fbiA*.

To assess the regulatory role of ResR, we used RNA sequencing to assess gene expression changes in the three independent *resR* mutants compared with isogenic wild-type *Mtb*. The *resR* mutants had ~2-fold higher levels of *whiB2* transcript but no significant alterations in the expression of the other genes adjacent to the major ResR-binding sites (Fig. 5C). The changes in *whiB2* expression were part of a pattern of altered transcription that also included increased expression of the *iniBAC* operon, which was previously shown to be expressed in response to cell wall–targeting antibiotics (*18*) and other mediators or regulators of central carbon metabolism (e.g., *bkdABC* and *clgR*) and cell wall–acting genes, notably *pbpB* (*19*, *20*) (fig. S11A). To further define the regulatory relationship between *resR* and *whiB2*, we reconstructed several *resR* variants in *Mycobacterium smegmatis* (homolog: *MSMEG_3644*), including CRISPR-i knockdown, overexpression, and point-mutant strains (Fig. 5D). Transcriptional profiling of these strains suggested that ResR functions as an activator of *whiB2* expression, and that the clinical mutation T77A results in increased *whiB2* transcription (Fig. 5D and fig. S11B).

WhiB2 is an essential regulator that controls cell growth division in mycobacteria. Knockdown of *whiB2* expression results in cell elongation and division defects, whereas overexpression of *whiB2* in *Mtb* results in increased cell length (*21*, *22*). We next sought to determine the effect of the *whiB2-fbiA* mutation on *whiB2* expression and *Mtb* physiology. We generated isogenic *Mtb* strains carrying a clinically prevalent mutation in *whiB2-fbiA* (3640375 C>T). The point-mutant strain had an ~1.6-fold increase in *whiB2* expression (Fig. 5E and fig. S12A). Moreover, this strain also phenocopied *resR* mutants, with similar morphotypic changes (longer and wider cells) (fig. S12B), a small increase of isoniazid MIC, and, most importantly (fig. S12C), faster recovery after antibiotic exposure (fig. S12, D and E). However, the magnitude of the resilience phenotype in the *whiB2-fbiA* mutant appeared to be smaller than that of *resR* mutants, suggesting that additional, as yet undefined, mechanisms might be involved.

resR, *whiB2-fbiA*, and *whiA* mutants are associated with acquisition of canonical drug resistance and treatment failure

The WhiB2 regulon has been best characterized in *Streptomyces*, in which the WhiB2 homolog

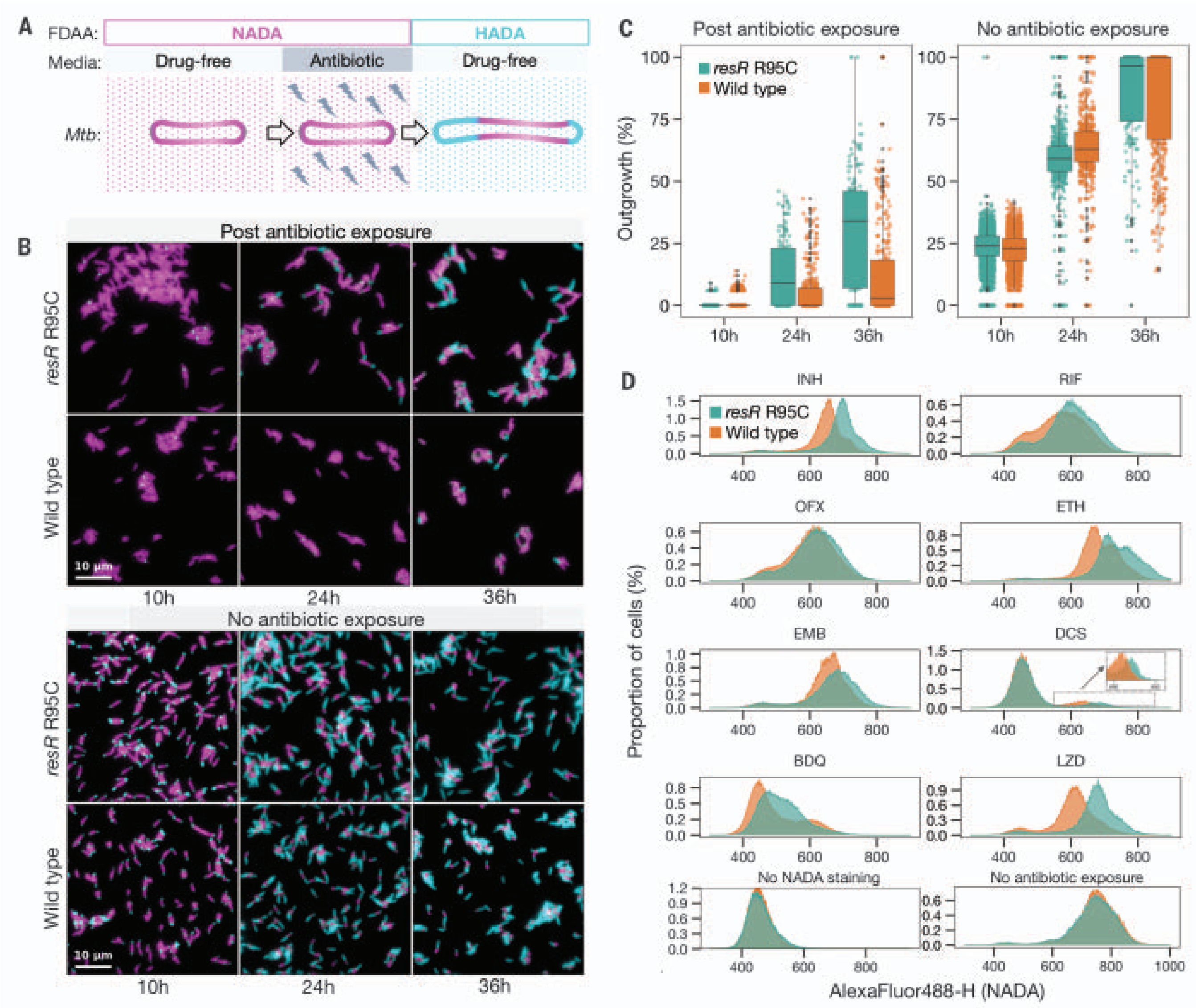


Fig. 4. Antibiotic resilience characterized by more rapid resumption of cell wall synthesis. (**A**) Schematic diagram of a pulse-chase experiment for tracking the regrowth of *Mtb* cells after antibiotic exposure using FDAA incorporation (NADA and HADA). (**B**) Representative microscopy snapshots of the *resR* mutants and wild type during regrowth after 24 hours of INH (100-fold MIC) exposure (top) and no antibiotic exposure (bottom). (**C**) Quantitative comparison of outgrowth between *resR* mutant and wild type in the post-antibiotic-exposure group and the no-antibiotic-exposure group; $P < 0.0001$ for 24 and 36 hours of post-antibiotic-exposure group by double-sided Kolmogorov–Smirnov test. (**D**) Flow cytometry for NADA incorporation into *Mtb* cells after 24 hours of recovery with or without antibiotic exposure.

functions with another regulator, WhiA, to initiate cell division (*23*, *24*). We found in our population genomic analysis that *whiA*, like *resR* and *whiB2*, is also a major target of positive selection in clinical *Mtb* isolates (Fig. 1B). To assess the clinical relevance of mutations in this regulatory triad, we compared the prevalence of fixed mutations of *resR, whiB2-fbiA*, and *whiA* in drug-susceptible strains versus drug-resistant strains. Of the 51,229 genomes in our database, 6.9% had mutations in these genes or IGRs, and mutations in these genes were significantly enriched in drug-resistant strains (fig. S13A). We further looked at *Mtb* strains from two high-TB-burden countries, India and China, where *Mtb* populations were not skewed by recent expansions of clonal *Mtb* strains, and each had >1000 sequenced isolates. Again, strains with mutations in the three genes or IGR had increased odds ratios (ORs) for drug resistance: India, OR: 1.80; 95% confidence interval (CI): 1.11~2.91; China, OR: 1.71; 95% CI: 1.06~2.74 (Fig. 6A and fig. S13B). Therefore, these data suggest that *resR*, *whiB2-fbiA*, and *whiA* mutations facilitate the evolution of canonical drug resistance. These mutations remain under positive selection in highly drug-resistant strains, as assessed by within-host evolution (Fig. 6B), suggesting that even after the emergence of resistance to first-line drugs, they provide an advantage to the bacterium as it encounters second-line agents. Evidence of selection at every level of drug resistance further supports the model that these variants are selected by a multidrug phenotype such as antibiotic resilience. However, the shifted isoniazid MIC is likely to contribute to their success in drug-susceptible strains, and this dual benefit may be one reason that they are clinically prevalent.

Finally, although *resR*, *whiB2-fbiA*, and *whiA* variants are more frequent in drug-resistant strains, variations in these genes are also under positive selection in drug-susceptible strains (fig. S13C), and the sites of mutation

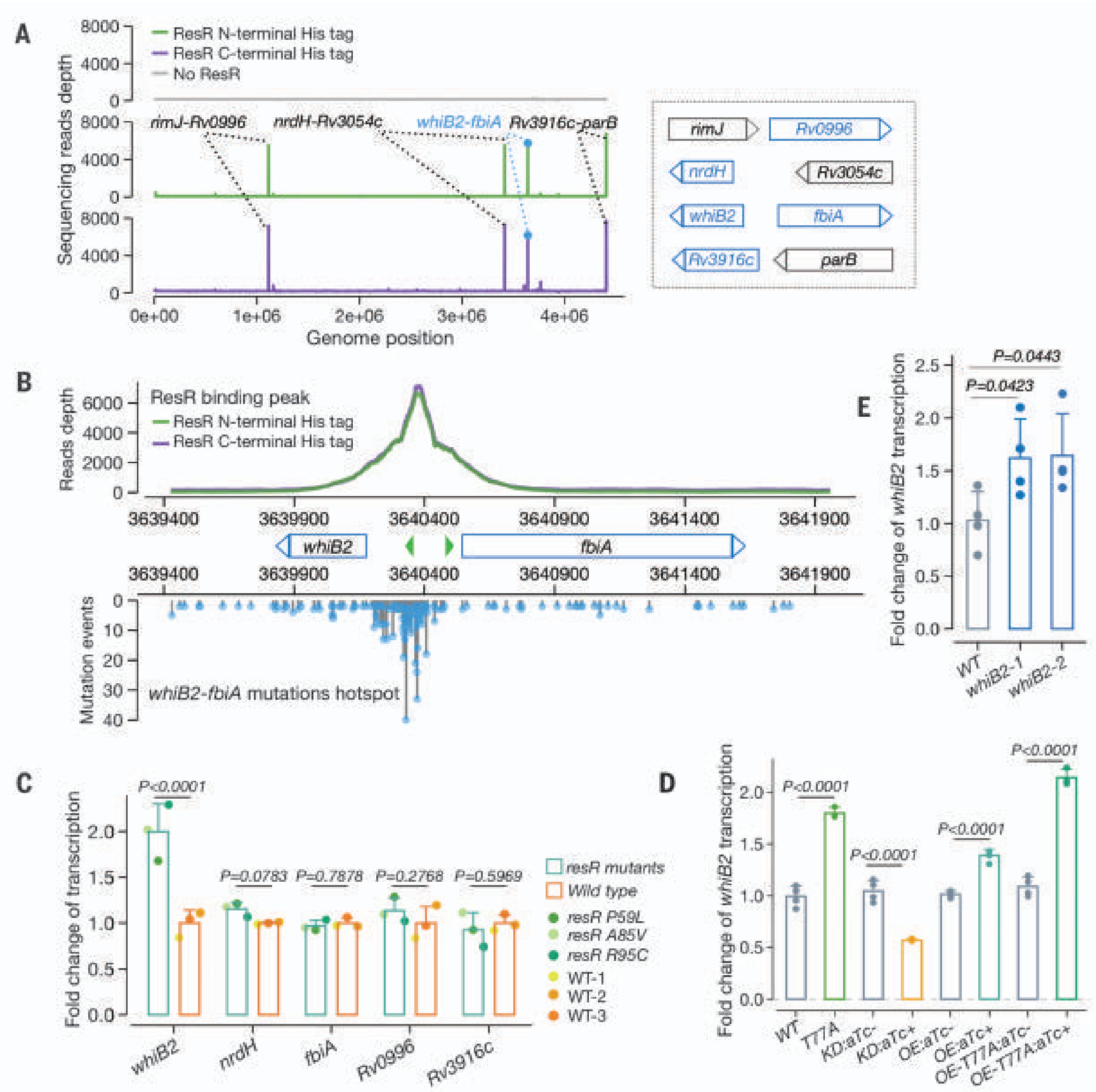


Fig. 5. ResR activates *whiB2* and clinical mutations leading to up-regulation of *whiB2*. (**A**) IDAP-Seq identified binding sites of ResR. (**B**) ResR-binding peaks overlaid with the mutations identified in clinical strains between *whiB2* and *fbiA*. The transcriptional start sites of *whiB2* and *fbiA* are indicated by green arrows. (**C**) Fold change of transcription of putative ResR targets in *resR* mutants and wild-type strains. *P* values were determined by the Wald test implemented in DEseq2. (**D**) Transcriptional changes of *whiB2* in *M. smegmatis* strains: wild-type (WT), *resR* point mutant (T77A), CRISPR-i knock-down of *resR* (KD), merodiploid overexpression of wild-type *resR* (OE) or T77A mutant form (OE-T77A). The absence or presence of aTc is indicated by (−) or (+). *P* values were determined by unpaired *t* test. (**E**) The *whiB2-fbiA* mutant (3640375 C>T) exhibited up-regulation of *whiB2*. *whiB2-1* and *whiB2-2* refer to the two parallel mutants. *P* values were determined by unpaired *t* test.

in drug-susceptible strains mimic those in drug-resistant strains (fig. S13D). Indeed, 1.5 to 9.7% of drug-susceptible *Mtb* strains from the high-TB-burden countries had fixed mutations in one of these three genes or IGRs (fig. S13E). We postulated that in addition to being clinically important as stepping-stone mutations for the emergence of canonical drug resistance, these variants independently contribute to treatment failure in drug-susceptible patients. To test this hypothesis, we reanalyzed data from the REMoxTB trial, a global phase 3 clinical trial that sought to reduce the treatment duration of drug-susceptible TB (*25*, *26*). Of the 1931 patients enrolled in the trial, paired *Mtb* isolates from 36 recurrent TB patients were whole-genome sequenced and the isolates remained drug sensitive (*26*). We found that eight of 36 (22.2%) isolates from patients who failed treatment had *resR*, *whiB2-fbiA*, or *whiA* mutations (Fig. 6C), a frequency significantly higher than found in the background populations of South Africa (108/2998, 3.6%) and Thailand (110/2036, 5.4%), where the trial was conducted (P = 0.0024 and P = 0.0152, respectively, Fisher's exact test; Fig. 6D). Additionally, the within-host frequency of these variants increased between the initial and recurrent episodes, in three cases rising to fixation (Fig. 6E).

We have identified a new form of altered drug susceptibility that we call antibiotic resilience, which is regulated by the gene *resR*. Mutations in the regulating cascade are positively selected in clinical *Mtb* isolates and associated with poor treatment outcomes in TB. Current

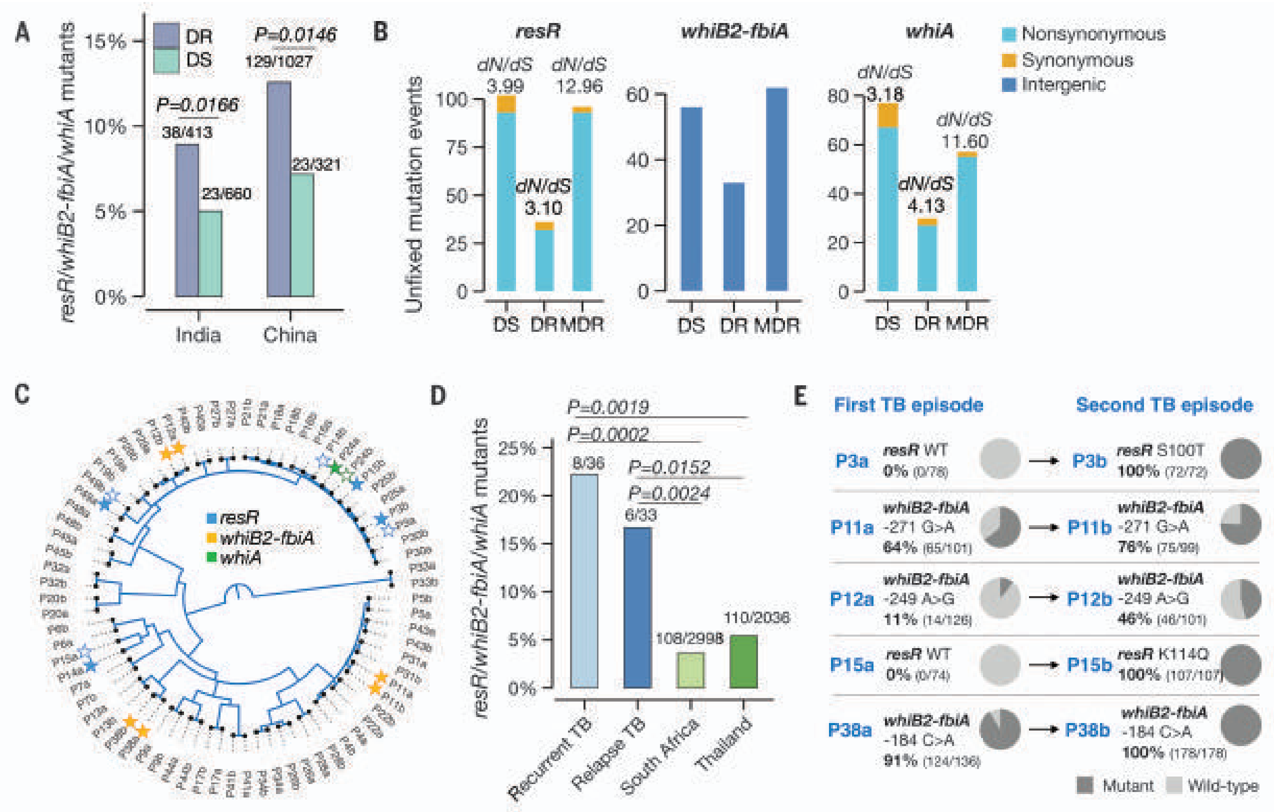


Fig. 6. *resR*, *whiB2-fbiA*, and *whiA* mutants are associated with canonical drug resistance and relapse of drug-susceptible TB. (**A**) Proportion of *resR*, *whiB2-fbiA*, and *whiA* mutants in DR and DS strains sequenced from India and China. *P* values were determined by Fisher's exact test. (**B**) Unfixed mutations in DS, DR, and MDR (resistant to RIF and INH) strains. (**C**) Phylogenetic tree of paired *Mtb* isolates from 36 recurrent TB patients. Solid stars indicate isolates in which mutations were detected, and empty stars indicate the absence of mutations in one of the paired isolates. (**D**) Percentage of isolates with mutations in *resR/whiB2-fbiA/whiA*. "Recurrent TB" includes three patients with *Mtb* isolates suggestive of reinfection. (**E**) Mutational trajectory in *Mtb* isolates from the first TB to second TB episodes in five pairs of isolates. The mutations in other three pairs were P14 (*resR*, D144N; 100% in P14a, 0% in P14b), P49 (*resR*, R95H; 8.7% in P49a, 0% in P49b), and P24 (*whiA*, A131T; 11.6% in P24a, 0% in P24b).

clinical measures of antibiotic susceptibility focus largely on the magnitude of drug exposure; that is, what concentration of drug the pathogen can experience and grow or at least survive (*27*). However, our results suggest that in the clinical setting, *Mtb* is evolving to change the temporal dynamics of its recovery after drug exposure. The temporal dynamics of drug responses are considered when developing new drugs and drug regimens but are often forgotten as potential drivers of treatment outcomes that might be influenced by pathogen variation. In studies of *Escherichia coli* experimentally evolved under intermittent drug exposure, temporal variants were the first to emerge (*28*). Further, the mutants had delayed regrowth rather than faster regrowth, suggesting that the temporal phenotypes could vary by the pathogen, drug regimen, and host factors such as drug metabolism (*29*–*31*). Because current clinical diagnostics are largely blind to these phenotypes, selection for temporal phenotypes such as antibiotic resilience may underlie treatment failure in other drug-susceptible infections.

REFERENCES AND NOTES

1. R. A. Fisher, B. Gollan, S. Helaine, *Nat. Rev. Microbiol.* **15**, 453–464 (2017).
2. J. Liu, O. Gefen, I. Ronin, M. Bar-Meir, N. Q. Balaban, *Science* **367**, 200–204 (2020).
3. A. Harms, E. Maisonneuve, K. Gerdes, *Science* **354**, aaf4268 (2016).
4. H. Safi *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **116**, 19665–19674 (2019).
5. J. F. Murray, D. E. Schraufnagel, P. C. Hopewell, *Ann. Am. Thorac. Soc.* **12**, 1749–1759 (2015).
6. Q. Liu *et al*., *Sci. Adv.* **6**, eaba4901 (2020).
7. A. Trauner *et al*., *Genome Biol.* **18**, 71 (2017).
8. Á. Chiner-Oms, M. G. López, M. Moreno-Molina, V. Furió, I. Comas, *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2113600119 (2022).
9. World Health Organization, *Technical Manual for Drug Susceptibility Testing of Medicines Used in the Treatment of Tuberculosis* (WHO, 2018).
10. N. D. Hicks *et al*., *Nat. Microbiol.* **3**, 1032–1042 (2018).
11. I. Levin-Reisman *et al*., *Nat. Methods* **7**, 737–739 (2010).
12. G. G. Zhanel, D. J. Hoban, G. K. Harding, *DICP* **25**, 153–163 (1991).
13. W. H. Beggs, J. W. Jenne, *Tubercle* **50**, 377–385 (1969).
14. J. K. Srimani, S. Huang, A. J. Lopatkin, L. You, *Mol. Syst. Biol.* **13**, 948 (2017).
15. K. J. Kieser, E. J. Rubin, *Nat. Rev. Microbiol.* **12**, 550–562 (2014).
16. M. T. M. Hannebelle *et al*., *Nat. Commun.* **11**, 452 (2020).
17. N. D. Hicks *et al*., *PLOS Pathog.* **16**, e1009063 (2020).
18. D. Alland, A. J. Steyn, T. Weisbrod, K. Aldrich, W. R. Jacobs Jr., *J. Bacteriol.* **182**, 1802–1811 (2000).
19. J. C. Betts, P. T. Lukey, L. C. Robb, R. A. McAdam, K. Duncan, *Mol. Microbiol.* **43**, 717–731 (2002).
20. G. Kumar, C. Galanis, H. R. Batchelder, C. A. Townsend, G. Lamichhane, *MSphere* **7**, e0003922 (2022).
21. J. E. Gomez, W. R. Bishai, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 8554–8559 (2000).
22. T. R. Raghunand, W. R. Bishai, *Microbiology (Reading)* **152**, 2735–2747 (2006).
23. M. J. Bush, G. Chandra, M. J. Bibb, K. C. Findlay, M. J. Buttner, *mBio* **7**, e00523-16 (2016).
24. M. J. Bush, M. J. Bibb, G. Chandra, K. C. Findlay, M. J. Buttner, *mBio* **4**, e00684-13 (2013).
25. A. Jindani *et al*., *N. Engl. J. Med.* **371**, 1599–1608 (2014).
26. J. M. Bryant *et al*., *Lancet Respir. Med.* **1**, 786–792 (2013).
27. CRyPTIC Consortium and the 100,000 Genomes Project, *N. Engl. J. Med.* **379**, 1403–1415 (2018).
28. O. Fridman, A. Goldberg, I. Ronin, N. Shoresh, N. Q. Balaban, *Nature* **513**, 418–421 (2014).
29. N. Strydom *et al*., *PLOS Med.* **16**, e1002773 (2019).
30. A. A. Ordonez *et al*., *Nat. Med.* **26**, 529–534 (2020).
31. K. Dheda *et al*., *Am. J. Respir. Crit. Care Med.* **198**, 1208–1219 (2018).
32. Q. Liu, J. Zhu, Code for: Tuberculosis treatment failure associated with evolution of antibiotic resilience, Zenodo (2022).

ACKNOWLEDGMENTS

We thank T. M. Walker for collecting and sharing information on the geographic origin of >10,000 *Mtb* isolates and C. M. Sassetti for the helpful discussions during this work.

Funding: This work was supported by the National Institutes of Health (grants P01 AI132130 and RFA-AI-21-065 to S.M.F., grant P01 AI143575 to S.M.F and E.J.R., and NIH/NIAID grant R01 AI143611-01 to B.B.A.). **Author contributions:** Q.L., J.Z., C.L.D., S.S., S.W., E.S.C., X.W., P.C., Y.J.L., N.D.H., G.H.B., and S.R.G. performed experiments and analyzed data. B.B.A., E.C.G., E.J.R., M.C.C., and S.M.F supervised the study. Q.L., J.Z., M.C., and S.M.F wrote the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials. Analysis code is deposited at GitHub (https://github.com/MtbEvolution/resR_Project) and archived at Zenodo (32). Sequencing data are deposited in the Sequence Read Archive database (PRJNA820190). **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abq2787

Materials and Methods
Figs. S1 to S13
Table S1
References (*33–61*)
MDAR Reproducibility Checklist

Submitted 29 March 2022; accepted 10 November 2022
10.1126/science.abq2787

CRYSTALLIZATION

Liquid metal synthesis solvents for metallic crystals

Shuhada A. Idrus-Saidi[1]†, Jianbo Tang[1]*†, Stephanie Lambie[2], Jialuo Han[1], Mohannad Mayyas[1], Mohammad B. Ghasemian[1], Francois-Marie Allioux[1], Shengxiang Cai[1], Pramod Koshy[3], Peyman Mostaghimi[4], Krista G. Steenbergen[5], Amanda S. Barnard[6], Torben Daeneke[7]*, Nicola Gaston[2]*, Kourosh Kalantar-Zadeh[1,8]*

In nature, snowflake ice crystals arrange themselves into diverse symmetrical six-sided structures. We show an analogy of this when zinc (Zn) dissolves and crystallizes in liquid gallium (Ga). The low-melting-temperature Ga is used as a "metallic solvent" to synthesize a range of flake-like Zn crystals. We extract these metallic crystals from the liquid metal solvent by reducing its surface tension using a combination of electrocapillary modulation and vacuum filtration. The liquid metal–grown crystals feature high morphological diversity and persistent symmetry. The concept is expanded to other single and binary metal solutes and Ga-based solvents, with the growth mechanisms elucidated through ab initio simulation of interfacial stability. This strategy offers general routes for creating highly crystalline, shape-controlled metallic or multimetallic fine structures from liquid metal solvents.

Perhaps the most captivating example of crystal growth is the formation of six-branched ice crystals (snowflakes) (*1*, *2*). Although they are distinct from one another, snowflakes constitute highly symmetrical six-branched dendrites, plates, prisms, and needle structures. The establishment of snowflake crystals relies on the orderly buildup of water molecules from an oversaturated gas phase onto the crystal facets dominated by molecular interactions. The synthesis of artificial snowflakes has allowed a phenomenological understanding of the crystallization process (*1*, *2*). A snowflake's sixfold symmetry stems from the hexagonal arrangement of the ice lattice as well as ambient growth conditions, including temperature and humidity, which in turn determine the crystals' morphologies (*3*). The morphological evolution of snowflake crystals is just one example of the highly complex crystal growth behaviors found in many natural and synthetic systems (*4*–*7*).

Here, extending the analogy of ice snowflakes to metals, we show that metallic snowflakes can be grown in metallurgical processes using liquid metals. The features that distinguish a liquid metal solvent from its nonmetallic counterparts are intuitive, as it is a dense and strongly interacting liquid that can solvate metal atoms in their neutral state, thus avoiding any need for the reduction of a metal precursor. In principle, the relatively isotropic nature of bonding in metals would not suggest structural anisotropy induced by the liquid metal solvent. However, in the case of a low-melting-temperature p-block post-transition metal such as gallium (Ga), the tendency for lower-density structures to form in the solid state and the fact that covalent character coexists with metallic bonding make the assumption of isotropic interactions overly simplistic (*8*). Given the considerable evidence of local structuring within liquid Ga at low temperatures (*9*–*11*), we hypothesize that liquid Ga can preferentially direct the growth of solute metal structures through a preference for either symmetry or orientation during crystal growth after nucleation. Consequently, rich crystal formation dynamics and diversified crystal morphologies are expected to emerge from liquid alloy systems. However, extracting the synthesized crystals from the liquid metal solvent while preserving their microscopic fine features and revealing their exposed crystal facets poses a considerable challenge.

The emergence of Ga and Ga-based liquid alloys offers viable possibilities for growing metallic crystals. Given that, at moderate temperatures, most metals in the periodic table of elements have a certain solubility in Ga and its alloys (*12*–*14*), a great variety of crystals can potentially be synthesized via the proposed route. This route can also be extended to some semimetals, such as silicon, but not to most nonmetals, which have limited solubility in liquid metals at low temperatures. Although this concept can lead to great opportunities, accessing the synthesized crystals grown inside the liquid metal solvent remains a hurdle. This challenge may be related to the affinity (wettability) between the crystals and metallic solvents. Unlike conventional synthesis, Ga-based liquid metals are not volatile, and as such, their high surface tension needs to be broken to isolate the generated crystals from the solvent (*15*). Thus, conventional physical sieving and filtering that work for separating synthesized particles from aqueous and organic solvents cannot be applied to liquid metals. Given these challenges, the possibility of extracting submicrometer and nanodimensional crystals from inside a liquid metal reaction medium has been largely ignored. We hypothesize that one effective way to access such nanostructures is to direct the generated crystals inside a liquid metal toward its surface and then reduce its surface tension by electrocapillary modulation (*16*, *17*).

In this study, we explored the morphology and composition selection of the crystal formation process of metals dissolved in liquid Ga as a solvent. We demonstrated that the metal crystals can be effectively extracted from the liquid Ga solvent by combining electrocapillary modulation and vacuum filtration. We used the binary zinc-gallium (Zn-Ga) system as the main model that gives rise to various types of highly symmetrical Zn structures, resembling that of the six-branched snowflake crystals. The influence of solute concentration, crystal growth time, temperature, and

[1]School of Chemical Engineering, University of New South Wales (UNSW), Kensington, NSW 2052, Australia. [2]MacDiarmid Institute for Advanced Materials and Nanotechnology, Department of Physics, University of Auckland, Auckland 1010, New Zealand. [3]School of Materials Science and Engineering, University of New South Wales (UNSW), Kensington, NSW 2052, Australia. [4]School of Minerals and Energy Resources Engineering, University of New South Wales (UNSW), Kensington, NSW 2052, Australia. [5]MacDiarmid Institute for Advanced Materials and Nanotechnology, School of Chemical and Physical Sciences, Victoria University of Wellington, Wellington 6140, New Zealand. [6]School of Computing, Australian National University, Acton, ACT 2601, Australia. [7]School of Engineering, RMIT University, Melbourne, VIC 3000, Australia. [8]School of Chemical and Biomolecular Engineering, University of Sydney, Darlington, NSW 2008, Australia.
*Corresponding author. Email: jianbo.tang@unsw.edu.au (J.T.); torben.daeneke@rmit.edu.au (T.D.); n.gaston@auckland.ac.nz (N.G.); k.kalantar-zadeh@unsw.edu.au (K.K.-Z.)
†These authors contributed equally to this work.

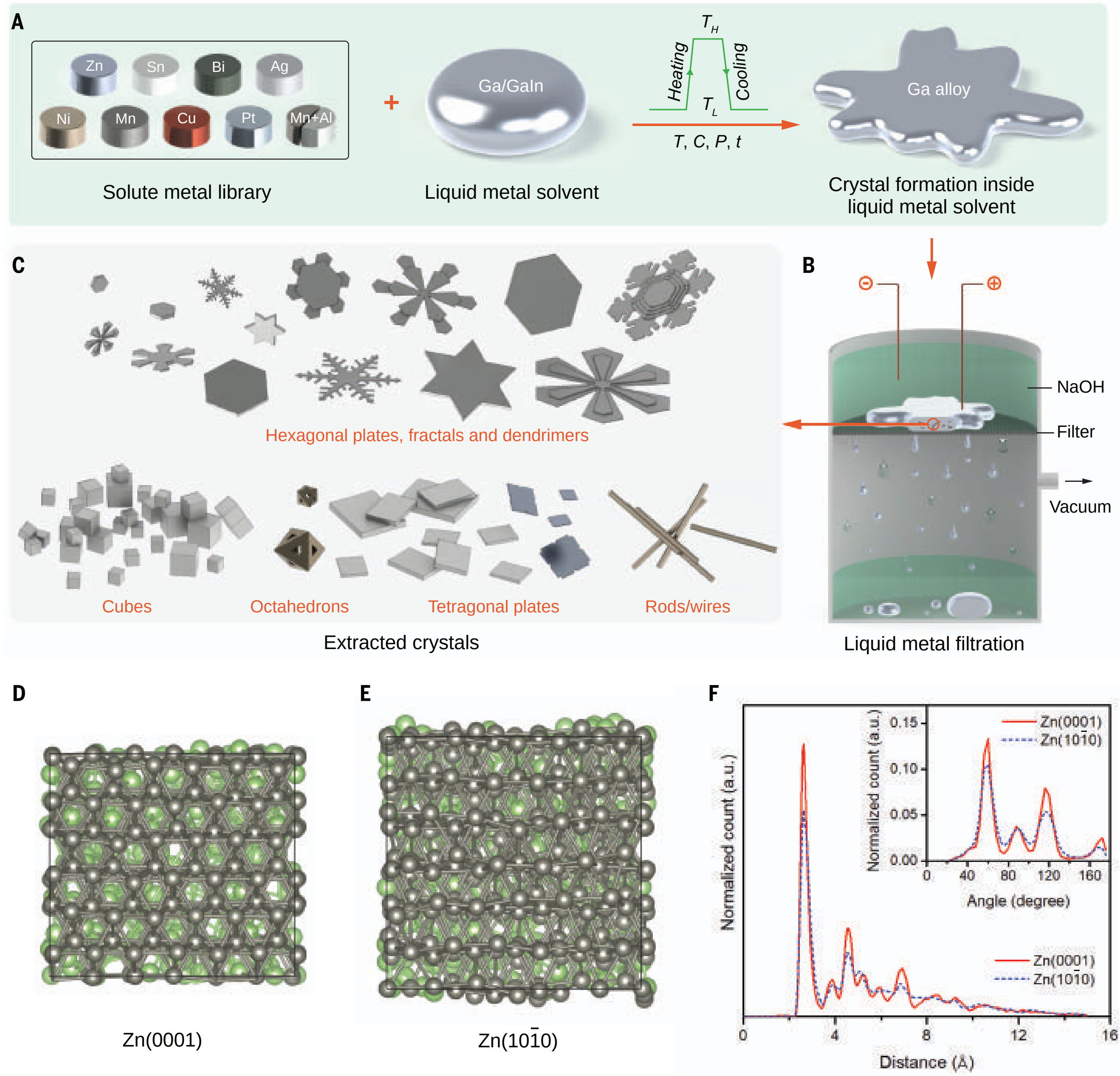


Fig. 1. Schematic presentation of the crystal formation and extraction strategy. (**A**) Experimental procedures to grow metal crystals in a liquid metal solvent. Solute metal, solvent metal, and the parameters (including temperature *T*, concentration *C*, pressure *P*, and time *t*) that determine the crystal structure and size are indicated. (**B**) Separating crystals from the liquid metal using an electrical field together with vacuum filtration. (**C**) Examples of various types of crystals featuring different morphologies and compositions. (**D** and **E**) Top-view snapshot structures of (D) the Zn(0001) facet and (E) the Zn(10$\bar{1}$0) facet interfaced with liquid Ga. Zn atoms are in dark gray, and the underlying Ga atoms are in green. (**F**) Radial distribution functional and angular distribution functional analyses for the Zn(0001) and Zn(10$\bar{1}$0) surface terminations when interfaced with liquid Ga. a.u., arbitrary units.

pressure were investigated for Zn-Ga systems and related to snowflake-like crystal morphologies. Subsequently, we tested other solute metals of tin (Sn), bismuth (Bi), silver (Ag), manganese (Mn), nickel (Ni), copper (Cu), platinum (Pt), the solute metal couple of Mn and aluminum (Al), and the solvent effect, to demonstrate the morphological and compositional diversity of crystals and showcase possible applications.

We synthesized a large variety of metallic crystals featuring distinct crystalline shapes using Ga liquid metal as the reaction medium (Fig. 1). Solute metals were first dissolved in Ga at an elevated temperature (T_H) (Fig. 1A and table S1). After homogeneous liquid alloys were formed, we cooled them to room temperature (T_L) to allow the solute metals to become oversaturated and precipitate out of the liquid metal solvent in crystalline forms, while the Ga bulk remained in its liquid state owing to supercooling (*18*). We first attempted to filter the metallic colloidal solution through a porous nylon filter membrane (pore size:

20 μm). However, the high surface tension of liquid Ga prevented the separation of the metallic crystal precipitates from the liquid metal solvent. To break the surface tension, we applied a potential of +5 V to the liquid metal in NaOH solution (1 mol liter^{-1}), while simultaneously applying vacuum-assisted filtration (Fig. 1B and fig. S1). Electrochemical actuation has been shown to enable the tuning of a liquid metal's surface tension from its natural value to near zero (*19–21*). After applying the voltage, the liquid metal flattened and expanded instantly because of the drastic decrease in its surface tension (*22*). After reduction of the surface tension, the liquid Ga was able to pass through the filter together with the NaOH solution. Our measurements indicated that nearly the entire volume of liquid Ga could pass through the porous membrane, leaving behind crystals topped with traces of Ga. After an additional washing step using NaOH to remove the remaining Ga in the filtered residues, we obtained clean metallic crystals (Fig. 1C). Depending on the choice of metallic solute, the intrinsic crystal lattice structure, and crystallization dynamics, we could produce intricate and distinct crystal morphologies by tuning the growth conditions.

To demonstrate the distinguishing features of a strongly interacting liquid metal synthesis solvent and to test our hypothesis on the solvent's role in directing anisotropic crystal growth, we performed ab initio molecular dynamics (AIMD) simulations to calculate the interfacial energies for the interaction between the low-index (0001) facet and $(10\bar{1}0)$ facet (Bravais-Miller index) of hexagonal Zn crystals and a Ga liquid, Ga(l) (*23*). Given that a larger magnitude interfacial energy indicates a stronger interaction between the interfaces, the results demonstrate that the $Zn(10\bar{1}0)$/Ga(l) interface interacts more strongly (−8.290 eV nm^{-2}) than the Zn(0001)/Ga(l) interface (−6.381 eV nm^{-2}) (table S2). Because of the increased strength of interaction between $Zn(10\bar{1}0)$ and Ga(l) [i.e., the increased effect of the Ga(l) surface disorder on the Zn crystal structure], the structural ordering within the interfaced $Zn(10\bar{1}0)$ slab is reduced compared with that of the Zn(0001) crystallographic orientation (Fig. 1, D and E, and fig. S2). By comparison, in the more weakly interacting Zn(0001)/Ga(l) interface, the Zn(0001) slab has a more clearly defined structure compared with the interfaced Zn $(10\bar{1}0)$, as reflected by the more well-defined peaks (narrower peaks and/or lower minima between peaks) in both radial distribution functional analysis and angular distribution functional analysis (Fig. 1F). This result agrees nicely with the surface energies for Zn, which indicate that the Zn(0001) facet is more stable than the $Zn(10\bar{1}0)$ facet (*24*). Therefore, Zn(0001) faceting is expected to be favored during the growth of Zn crystals in liquid Ga, as $Zn(10\bar{1}0)$ is destabilized by its stronger and more structurally disruptive interfacial interaction.

The preferred Zn(0001) faceting predicted by the AIMD simulations is validated by our experimental observations of Zn crystal growth in liquid Ga. When Zn is used as the solute, the crystals formed in the liquid Ga strongly resemble snowflake morphologies, particularly when considering the highly varied crystal types obtained from a single Zn_x-Ga_y system (x wt % of Zn and y wt % of Ga). The Zn crystals extracted from the same $Zn_{10}Ga_{90}$ sample have different structures, as shown in the scanning electron microscopy (SEM) images in Fig. 2, A to C, and varying the growth conditions (time, temperature, and pressure) gives rise to substantial structural changes (Fig. 2, A to G). At 10%, the Zn content is above its solubility limit of 3.64 wt % in liquid Ga at room temperature. The observations suggest that increasing the growth duration, in general, leads to larger crystals (Fig. 2F). Despite their distinct shapes, all the crystals feature sixfold symmetry, which stems from the hexagonal lattice structure of Zn, as shown in our x-ray diffractogram (Fig. 2H). In addition to the AIMD simulations, we validated the formation of hexagonal Zn crystals on the basis of the thermodynamic properties of the Zn-Ga system using FACTSage 8.0 (fig. S3). From the x-ray diffraction (XRD) and energy-dispersive spectroscopy (EDS) maps (Fig. 2, H and I), the extracted crystals are clean and consist mostly of Zn and negligible amounts of Ga. The pristine Zn crystals are naturally oxidized on the surface when exposed to ambient air after extraction (fig. S4).

The influences of the control parameters and solute-solvent energetics and interactions on the obtained crystals can also be interpreted by the classical theory of crystal formation (*25*). Multiple parameters, including temperature, concentration, pressure, and growth orientation (facet), can affect the formed crystals substantially through different pathways (*23*). For the Zn-Ga system, hexagonal Zn nuclei form first during cooling. The Zn crystals we obtained are grown by the gradual attachment of Zn atoms from the oversaturated liquid metal solvent to different facets. As evidenced by our simulations, the Zn(0001) basal facet (Fig. 2H) with well-defined structure and order is preferentially preserved in interaction with liquid Ga during crystal growth, leading to the observation of platelike snowflake structures (Fig. 2, A to E) and the substantially higher (002) XRD peak in comparison to the (100) peak (Miller index) (Fig. 2H and fig. S5) (*23*). Apart from that, effects such as secondary nucleation, facet instability, and difference in the growth rate on different prism facets [induced by the changes in repeatable step sites for atom attachment and other growth conditions (*26*, *27*)] further lead to more complex crystal structures, such as fractals and dendrimers, rather than simple hexagonal plates. In addition, splitting of crystal tips (corners), branched growth, and jagged edges resembling the Koch curve (*28*) are frequently observed in these crystals.

Among all the influential variables, increasing the temperature at which the alloy was prepared to 550°C (Fig. 2D) generated a broad range of Zn crystal sizes that are typically larger than that of the 350°C case (Fig. 2A). At this higher temperature, the occurrence of 12-branched Zn crystals becomes frequent, coexisting with much smaller six-branched ones (Fig. 2D and fig. S6). As in the formation of 12-branched snow crystals (*29*), 12-branched Zn crystals are considered as the concurrent growth of two overlapping hexagonal crystal seeds. Therefore, the emergence of 12-branched Zn crystals at higher initial temperature means that, apart from growth rate, temperature also affects the growth regime of the crystals owing to changes in their thermodynamic growth path. Zn crystals grown at elevated pressure (5 bar) feature simple fractal shapes, which are also larger in size than those grown under ambient pressure (Fig. 2E), presumably as a result of facilitated nucleation due to a decrease in surface tension of liquid Ga (*30*).

The influence of growth time and solute concentrations on the morphology and size of the Zn crystals is further investigated in detail, which allows us to construct a morphology diagram for the Zn-Ga system (Fig. 3A). We show examples of extracted crystals from different Zn concentrations for day 2 in Fig. 3B and other growth times in figs. S7 to S9 (see also table S3). We summarize the characteristic shapes and relative sizes of Zn crystals grown at day 1, day 2, and day 10 across the Zn concentration range from 5 to 20 wt % (350°C initial temperature and ambient pressure) in Fig. 3A. The transition in crystal morphologies is determined by the tendency of anisotropic growth (*31*), with the hexagonal plates and the dendrimers having the lowest and highest anisotropies, respectively.

The morphology diagram reveals the signature sixfold symmetry of all the Zn crystals, despite their highly intricate shapes, which is a clear indication of the hexagonal Zn lattice structures and the Zn(0001) faceting revealed by our AIMD simulations. The formation of dendrimers is more favorable than simple fractals and plates at low Zn concentrations (Zn_5Ga_{95} and $Zn_{10}Ga_{90}$, except for Zn_5Ga_{95} at day 10), whereas high Zn concentration ($Zn_{20}Ga_{80}$) predominantly leads to a combination of fractals and plates. In addition, we observed at high Zn concentration more fractals than hexagonal plates (fig. S9). We saw the growth of the Zn

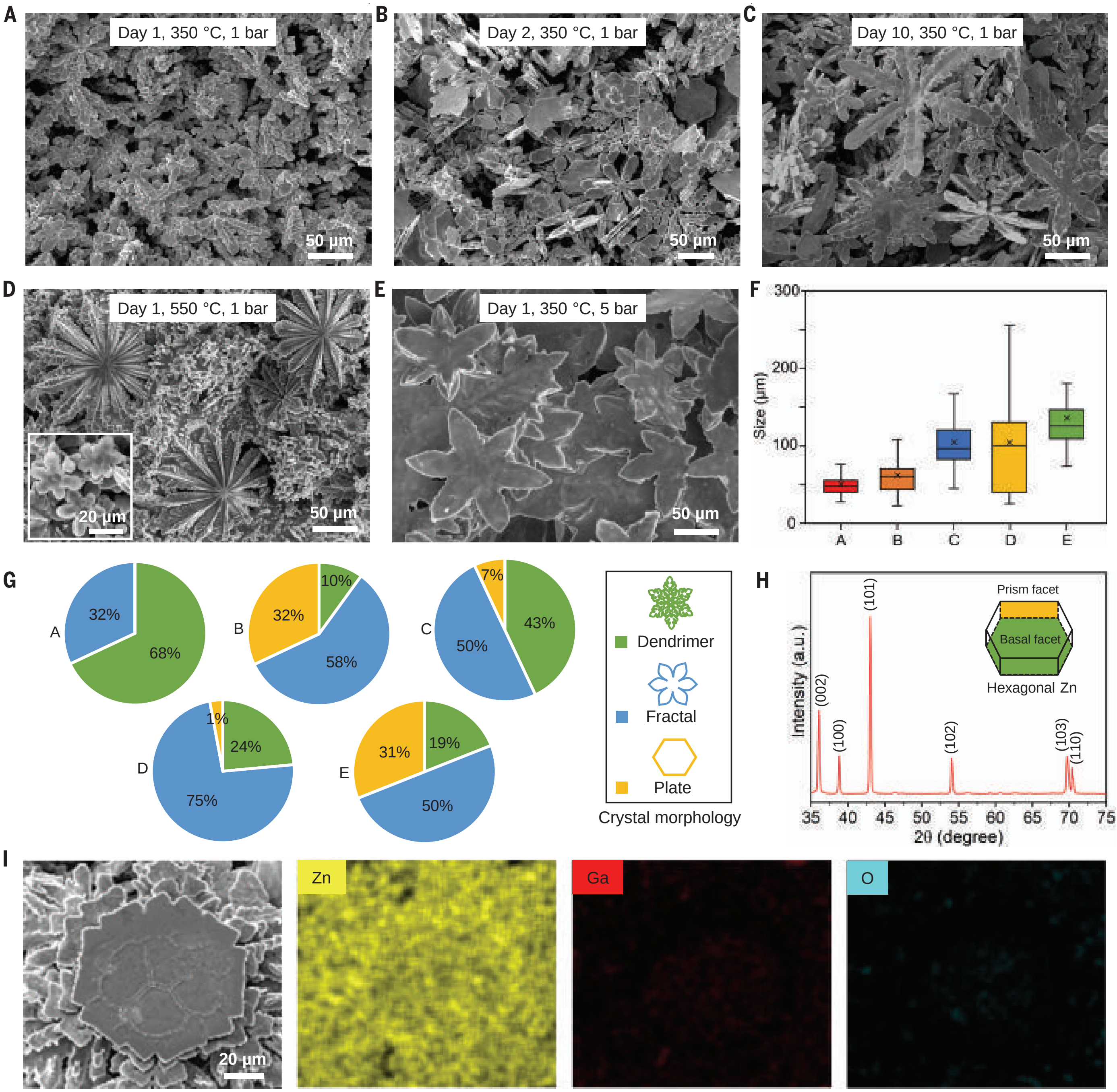


Fig. 2. Characterizations of Zn crystals extracted from liquid $Zn_{10}Ga_{90}$ alloy. (**A** to **C**) SEM images of Zn crystals obtained after (A) 1 day, (B) 2 days, and (C) 10 days of growth time at 350°C initial temperature. (**D**) SEM image of Zn crystals after 1 day of growth at 550°C initial temperature. (**E**) SEM image of Zn crystals formed after 1 day at 5 bar. (**F**) Size distribution of the crystals under different experimental conditions in (A) to (E). (**G**) The ratios of different crystal morphologies (dendrimers, fractals, and plates) in the extracted crystals of (A) to (E). (**H**) XRD patterns of the extracted crystals showing a single phase of Zn (Miller index). (**I**) Energy-dispersive x-ray spectroscopy elemental mapping showing the distribution of the Zn, Ga, and O.

crystals over time for the group with 10 wt % Zn. However, we did not find a size dependence on growth time of the crystals for the Zn_5Ga_{95} and $Zn_{20}Ga_{80}$ groups. Much larger dendrimers are obtained when the Zn concentration is further increased to 30 wt % after a day of growth time (fig. S10). The switching of growth regimes from dendrimers (Zn_5Ga_{95} and $Zn_{10}Ga_{90}$) to fractal and/or hexagonal plates ($Zn_{20}Ga_{80}$) and then back to dendrimers ($Zn_{30}Ga_{70}$) is reminiscent of the snow crystal morphology diagram (*2*). The observed trends suggest that metal crystal growth in a metallic liquid shares its highly dynamic nature with nonmetallic systems and is sensitive to the prevailing growth conditions. We used a machine learning model to confirm the sensitivity of the crystal structures (figs. S11 and S12). X-ray imaging (fig. S13) and SEM and EDS (fig. S14) were also performed to observe the crystal dimensions and morphologies at different growth times.

We further investigated the solvent effect by changing the liquid metal solvent from Ga to

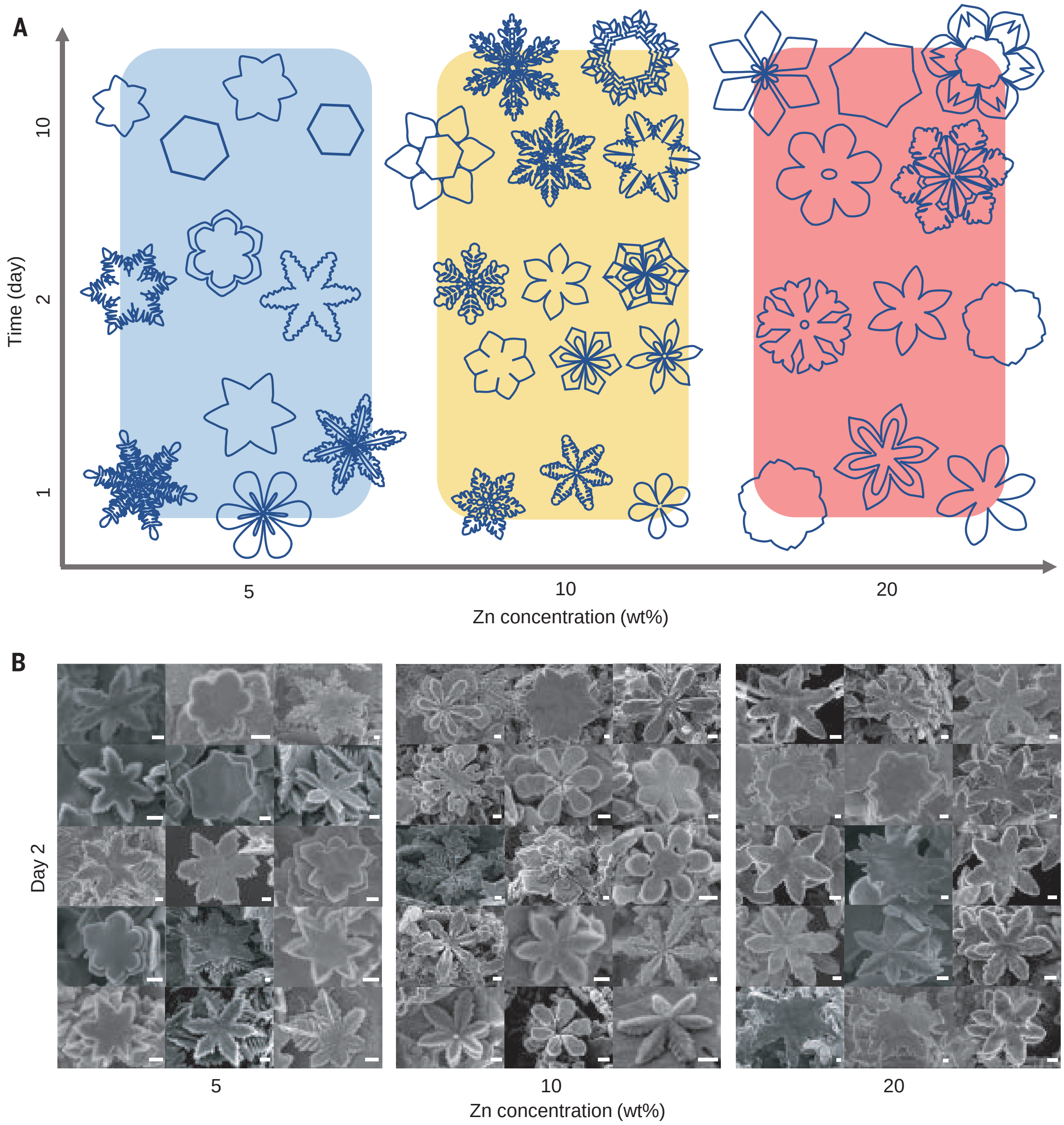


Fig. 3. Morphology diagram of the Zn crystals. (**A**) A morphology diagram drawn on the basis of the types of Zn crystals as a function of crystal growth time and Zn concentration under the same temperature (350°C) and pressure (atmospheric pressure) conditions. (**B**) Examples of Zn crystals with different Zn compositions after 2 days of growth time. Scale bars: 10 μm. See examples for different growth times in figs. S7 to S9.

Ga-indium (In) alloys for Zn crystal growth, which was found to generate the same crystal type (hexagonal Zn) but strengthen the (001) faceting (figs. S15 and S16). This is attributed to the decreased activation energy of the solute metal during nucleation as well as the reduced concentration barrier during crystal growth as a result of adding In to the Ga solvent (*23*), thereby leading to larger crystals at higher In ratios (fig. S17).

As a distinctive metallic solvent, liquid Ga dissolves most metals in the periodic table, albeit to different extents. Therefore, the method can be extended to many other solute metals for growing and extracting metallic crystals. We demonstrated this with the binary systems $Sn_{30}Ga_{70}$, Bi_2Ga_{98}, Ag_2Ga_{98}, Mn_2Ga_{98}, Ni_2Ga_{98}, Cu_2Ga_{98}, and Pt_2Ga_{98} (Fig. 4, A to G) and with the ternary $Al_5Mn_5Ga_{90}$ system (Fig. 4H). We chose the concentrations mainly according to their respective phase diagrams and solubility.

We obtained crystals of different morphologies and compositions depending on their intrinsic crystal structure and the influence of growth conditions on each individual system. The crystals grown in the binary systems can be classified into two categories. This first category includes Sn-Ga, Bi-Ga, and the Zn-Ga systems in which the solute metal phase completely separates from the liquid Ga phase during crystallization (no intermetallic phase), leading to single-metal crystals (Fig. 4, A and B). Predominantly nonfaceted dendritic growth of tetragonal Sn crystals reaching hundreds of micrometers in size is observed in the Sn-Ga system, whereas faceted growth of cubic Bi crystals emerges in the Bi-Ga system, generating exotic hollow hopper crystals that reach several micrometers in size. The other category, including Ag-Ga, Mn-Ga, Ni-Ga, Cu-Ga, and Pt-Ga, is distinguished by the formation of intermetallic compound phases between the solute metals and the solvent Ga. Both hexagonal Ag_2Ga and tetragonal MnGa intermetallic

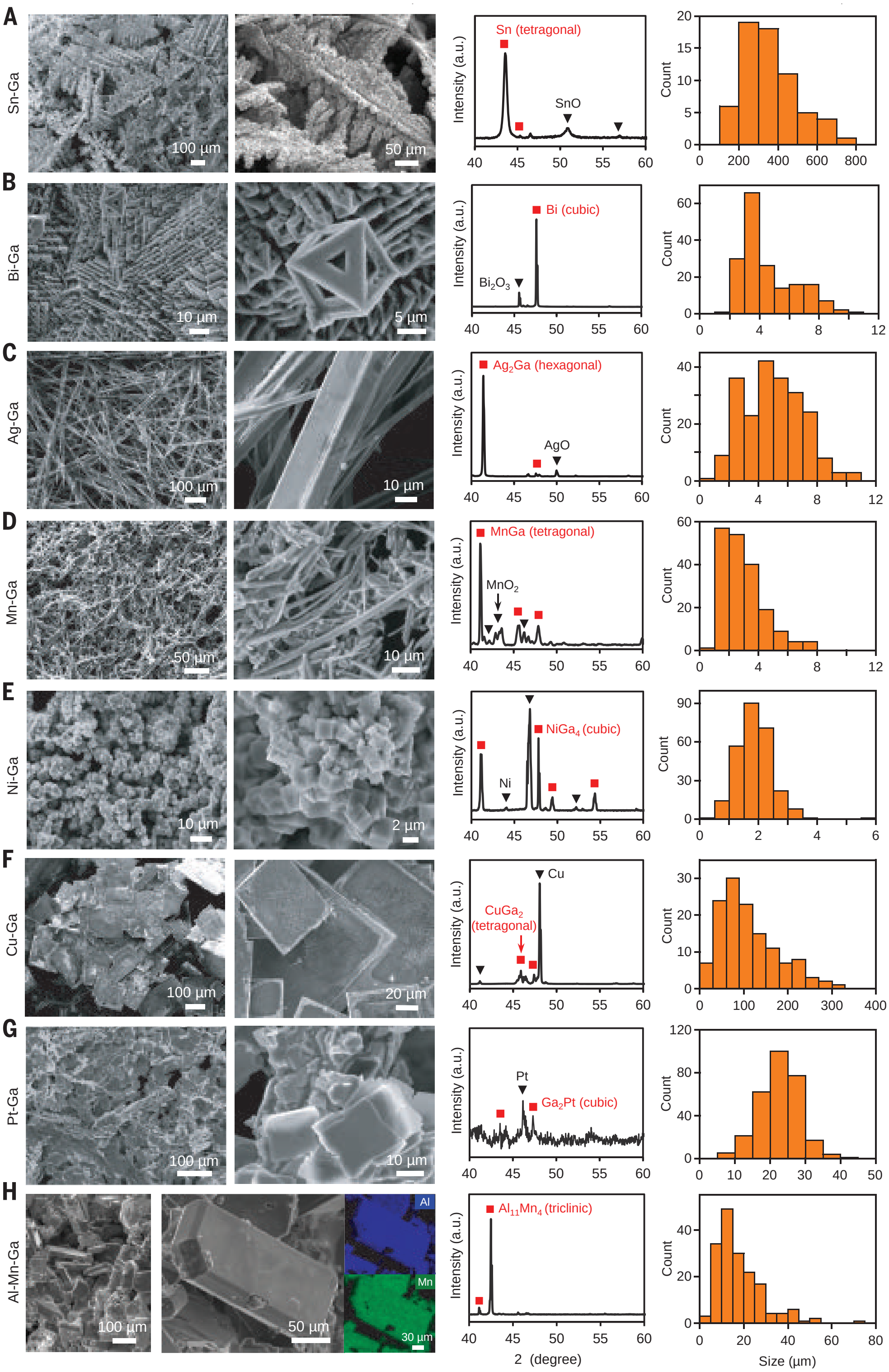


Fig. 4. Crystals formed and extracted from metal Ga systems other than Zn-Ga. (**A** to **G**) SEM images, XRD patterns and size distributions of the extracted crystals from the binary systems (A) Sn-Ga, (B) Bi-Ga, (C) Ag-Ga, (D) Mn-Ga, (E) Ni-Ga, (F) Cu-Ga, and (G) Pt-Ga (image on the right side shows both platelike and rodlike structures). (**H**) SEM image (right), EDS element mapping showing the distributions of Al and Mn, XRD pattern, and size distribution from a ternary Al-Mn-Ga alloy system. The growth conditions of all crystals are set to 1-day growth time, 350°C initial temperature, and atmospheric pressure. The oxide phases identified in the XRD of (A), (C), and (D) are formed after the crystals are extracted from liquid Ga and are not discussed. The size distributions are surveyed according to their (A) length of dendrites, [(B) and (G)] side length of the crystals, [(C) and (D)] thickness of the rod structures, (E) side length of the cubic structures, (F) diagonal length of the plates, and (H) thickness of the crystals.

crystals grow into high-aspect-ratio rod (or wire) structures with diameters that are typically less than 10 µm (Fig. 4, C and D), implying highly faceted and anisotropic growth of the two. The Ni-Ga, Cu-Ga, and Pt-Ga systems produce cubes of cubic $NiGa_4$ (Fig. 4E), plates of tetragonal $CuGa_2$ (Fig. 4F), and edge and corner truncated tetragonal plates of cubic Ga_2Pt crystals, together with larger rodlike and dendritic structures (Fig. 4G). Apart from the intermetallic phases, these three systems also show evidence of single-metal crystals of the solute metals. The elemental distributions of the metals in the crystals can be found in the EDS maps in fig. S18. The crystal phases we observed can be traced back to their phase diagram at the relevant concentration and temperature ranges (fig. S19).

Simultaneously introducing the two solute metals Al and Mn into liquid Ga leads to a ternary system. The crystal growth dynamics in a ternary system are expected to have greater complexity simply by virtue of the three-pair interactions between the metal

atoms. In comparison, a binary system has only one interacting metal atom pair. Surprisingly, as evidenced by the EDS elemental mapping and XRD data, the Al-Mn-Ga system produces a binary $Al_{11}Mn_4$ intermetallic phase that exclusively consists of the two solute metals (Fig. 4H). Given that the MnGa intermetallic phase is found to form in the binary Mn-Ga system (Fig. 4D), further introduction of Al must alter the energetics of crystal formation in the system. The $Al_{11}Mn_4$ has relatively low Gibbs energy of formation (*32*), and consequently, the emergence of $Al_{11}Mn_4$ under the current conditions should become energetically favored in comparison to any other possible intermetallic phases (fig. S20). The possibility, demonstrated here, of growing compound crystals through our simple yet efficient liquid metal solvent growth method greatly expands the accessible crystal library.

We further applied AIMD simulations to the Ag-Ga system to validate the role of interfacial energy and structure in directing the growth direction of liquid metal–grown crystals. Whereas the Ag_2Ga crystals grown from the Ag-Ga system share the same hexagonal lattice structure with the Zn crystals from the Zn-Ga system, Ag_2Ga forms long hexagonal rodlike structures (Fig. 4C) that are starkly different from the platelike hexagonal Zn flakes (Figs. 2 and 3). Such a distinct structural development of crystals grown from the two systems is indeed reflected in our simulations. The $Ag_2Ga(0001)$/Ga(l) interfacial interaction is increased (i.e., stronger) compared with that of the $Ag_2Ga(10\bar{1}0)$/Ga(l) interface (table S2). Correspondingly, the $Ag_2Ga(0001)$ termination, which is more strongly interacting with liquid Ga, is found to be less structured than the $Ag_2Ga(10\bar{1}0)$ surface termination (fig. S21). Using the same arguments as outlined for the Zn-Ga system, we determine that the reduced interaction between the $Ag_2Ga(10\bar{1}0)$ termination and liquid Ga allows a greater degree of structural integrity to be maintained in this surface termination than in the $Ag_2Ga(0001)$ termination. There will be less disruption of the $Ag_2Ga(10\bar{1}0)$ faceting by the disordered Ga(l), leading to the rod structures. The Sn-Ga system that showed dendritic (nonfaceted) Sn crystal growth in experiments was additionally investigated through AIMD simulations of interfacial energies. In this case, the Sn(200)/Ga(l) interface was found to be structurally stable, whereas the Sn(211)/Ga(l) interface was not—in essence, with the strength of the interfacial interaction being sufficient to fully disorder the Sn in this orientation (table S2). This strong destabilization is consistent with the dendritic growth of the Sn crystals observed in experiment (Fig. 4A). The comparison of the AIMD interfacial energetics and the structural characteristics of these selected systems explains the experimentally observed distinct crystal morphologies.

Here, we presented the synthesis of a variety of single metal and intermetallic crystals of different structures using liquid metals by careful selection of the solute and solvent metals, as well as by regulating the parameters that influence the growth dynamics. Particularly, it was possible to either adjust the growth dynamics without changing the crystal type (compare the Zn-Ga system in Figs. 2 and 3 and Zn-In-Ga system in figs. S15 to S17) or alter the generated crystals by introducing a suitable element (compare the Mn-Ga system in Fig. 4D and Mn-Al-Ga system in Fig. 4H). Finally, the obtained materials, with facet-dependent response to external fields or stimuli, enabled a range of applications for which proof-of-concept examples in gas sensing, piezoelectricity, and photocatalysis were demonstrated (figs. S22 to S24) (*23*).

We demonstrated that a variety of single metallic or bimetallic crystals can be grown in a liquid metal solvent to successfully extract crystals with fine features from the solvent using a combination of electrocapillary modulation and vacuum filtration sieving. The influences of solute concentration, solvent type, growth time, temperature, pressure, and the lattice structure of the nuclei on crystal growth were investigated to understand the growth dynamics. The morphology and size of the isolated crystals were rationalized by considering the unit crystal structure, effects of facet selecting and anisotropic growth, and growth rate. We presented the uncanny similarity of Zn metal flakes to ice snowflakes. In agreement with experimental observations, our simulations suggest that for a solid crystal of defined shape to grow within a liquid Ga solvent, at the nanoscale, it should be able to preserve its structure despite the presence of a strongly interacting liquid metal environment. The findings advance the understanding of crystal growth using liquid metals as the solvent and present a viable method for extracting these crystals. The use of liquid metals offers a previously unexplored pathway toward the growth of metallic nanostructures that does not rely on reactive soluble precursors, enabling unimpeded growth and leading to the formation of complex solidification patterns. By demonstrating both the crystal growth and extraction methods of metallic crystals, our work advances the field of functional nanoparticle synthesis while also offering opportunities in additive manufacturing and metallurgy for future manufacturing.

REFERENCES AND NOTES

1. U. Nakaya, *Snow Crystals: Natural and Artificial* (Harvard Univ. Press, 1954).
2. K. G. Libbrecht, *Rep. Prog. Phys.* **68**, 855–895 (2005).
3. J. Hallett, B. J. Mason, *Proc. R. Soc. London Ser. A* **247**, 440–453 (1958).
4. Y. Xia, Y. Xiong, B. Lim, S. E. Skrabalak, *Angew. Chem. Int. Ed.* **48**, 60–103 (2009).
5. M. S. Titus *et al*., *Sci. Adv.* **2**, e1601796 (2016).
6. A. M. Guloy *et al*., *Nature* **443**, 320–323 (2006).
7. M. G. Kanatzidis, R. Pöttgen, W. Jeitschko, *Angew. Chem. Int. Ed.* **44**, 6996–7023 (2005).
8. X. G. Gong, G. L. Chiarotti, M. Parrinello, E. Tosatti, *Phys. Rev. B* **43**, 14277–14280 (1991).
9. S.-F. Tsay, S. Wang, *Phys. Rev. B* **50**, 108–112 (1994).
10. L. E. González, D. J. González, *Phys. Rev. B* **77**, 064202 (2008).
11. K. G. Steenbergen, N. Gaston, *Nano Lett.* **16**, 21–26 (2016).
12. H. Okamoto, *Desk Handbook: Phase Diagrams for Binary Alloys* (ASM International, ed. 2, 2000).
13. J. Tang *et al*., *ACS Appl. Mater. Interfaces* **9**, 35977–35987 (2017).
14. J. Tang *et al*., *Nat. Nanotechnol.* **16**, 431–439 (2021).
15. X. Zhao, S. Xu, J. Liu, *Front. Energy* **11**, 535–567 (2017).
16. M. R. Khan, C. B. Eaker, E. F. Bowden, M. D. Dickey, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 14047–14051 (2014).
17. J. Zhang, L. Sheng, J. Liu, *Sci. Rep.* **4**, 7116 (2014).
18. J. Tang *et al*., *Nat. Synth.* **1**, 158–169 (2022).
19. M. Mayyas *et al*., *ACS Nano* **14**, 14070–14079 (2020).
20. C. B. Eaker, D. C. Hight, J. D. O'Regan, M. D. Dickey, K. E. Daniels, *Phys. Rev. Lett.* **119**, 174502 (2017).
21. M. D. Dickey, *ACS Appl. Mater. Interfaces* **6**, 18369–18379 (2014).
22. F. F. Yun *et al*., *Natl. Sci. Rev.* **7**, 366–372 (2020).
23. Materials and methods are available as supplementary materials.
24. R. Tran *et al*., *Sci. Data* **3**, 160080 (2016).
25. D. Turnbull, *J. Appl. Phys.* **21**, 1022–1028 (1950).
26. J. S. Langer, *Rev. Mod. Phys.* **52**, 1–28 (1980).
27. T. Melia, W. Moffitt, *Ind. Eng. Chem. Fundam.* **3**, 313–317 (1964).
28. L. M. A. Monzon, L. Klodt, J. M. D. Coey, *J. Phys. Chem. C* **116**, 18308–18317 (2012).
29. T. Kobayashi, Y. Furukawa, *J. Cryst. Growth* **28**, 21–28 (1975).
30. J. J. Sobczak, L. Drenchev, R. Asthana, *Int. J. Cast. Met. Res.* **25**, 1–14 (2012).
31. C.-L. Lu *et al*., *NPG Asia Mater.* **6**, e135 (2014).
32. Z. Yu *et al*., *Mater. Sci. Technol.* **33**, 2086–2096 (2017).

ACKNOWLEDGMENTS

Funding: The authors acknowledge support from an Australian Research Council (ARC) Laureate Fellowship grant (FL180100053), the ARC Center of Excellence FLEET (CE170100039), an ARC Discovery Early Career Researcher Award (DE220100816), and an ARC Discovery Project (DP220101923), the Marsden Fund Council from Government funding, managed by Royal Society Te Apārangi. The use of New Zealand eScience Infrastructure (NeSI) high performance computing facilities and University of New South Wales Tyree Micro CT Imaging Facility is also acknowledged. **Author contributions:** Conceptualization: S.A.I.-S., J.T., T.D., and K.K.-Z. Data curation: S.A.I.-S. and J.T. Methodology: S.A.I.-S., J.H., M.M., M.B.G., F.-M.A., S.C., P.K., P.M., and A.S.B. Investigation: S.A.I.-S. and J.T. Formal analysis: S.A.I.-S. and J.T. Software: P.K. and A.S.B. MD simulation: S.L., K.G.S., and N.G. Funding acquisition: J.T., T.D., and K.K.-Z. Project administration: K.K.-Z. Supervision: J.T., T.D., and K.K.-Z. Writing – original draft: S.A.I.-S. and J.T. Writing – review & editing: J.T., T.D., and K.K.-Z. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials.

License information: Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abm2731
Materials, Methods, and Discussions
Figs. S1 to S24
Tables S1 to S3
References (*33*–*57*)

Submitted 6 September 2021; accepted 11 November 2022
10.1126/science.abm2731

PIEZOELECTRICS

Giant electric field–induced strain in lead-free piezoceramics

Geng Huangfu[1]†, Kun Zeng[2,3]†, Binquan Wang[1], Jie Wang[1], Zhengqian Fu[2]*, Fangfang Xu[2,3], Shujun Zhang[4]*, Haosu Luo[5], Dwight Viehland[6], Yiping Guo[1]*

Piezoelectric actuators are indispensable over a wide range of industries for their fast response and precise displacement. Most commercial piezoelectric actuators contain lead, posing environmental challenges. We show that a giant strain (1.05%) and a large-signal piezoelectric strain coefficient (2100 picometer/volt) are achieved in strontium (Sr)–doped (K,Na)NbO_3 lead-free piezoceramics, being synthesized by the conventional solid-state reaction method without any post treatment. The underlying mechanism responsible for the ultrahigh electrostrain is the interaction between defect dipoles and domain switching. The fatigue resistance, thermal stability, and strain value (0.25%) at 20 kilovolt/centimeter are comparable with or better than those of commercial Pb(Zr,Ti)O_3-based ceramics, showing great potential for practical applications. This material may provide a lead-free alternative with a simple composition for piezoelectric actuators and a paradigm for the design of high-performance piezoelectrics.

Piezoelectric actuators can directly convert electrical signal into mechanical strain and have been widely applied in consumer electronic products, transportation, precise optical instruments, and microelectromechanical systems and robotics, to name a few (*1*, *2*). The world market for piezoelectric devices, in which the piezoelectric actuator segment holds the major share, is expected to grow up to US$35.4 billion by 2026.(*3*). For lead-free devices especially, the compound annual growth rate during 2019 to 2024 is expected to be as high as 20.8% (*4*). The highest electric field–induced strain among piezoelectric ceramics is 1.3%, reported in the lead-containing material (Bi,La)FeO_3–$PbTiO_3$, which is driven by a high electric field of 80 kV/cm (*5*). Considering environmental regulations against the use of lead, substantial efforts have been made to search for lead-free alternatives (*6*–*8*). Among the lead-free piezoelectric systems, $(Na_{1/2}Bi_{1/2})TiO_3$ (NBT)–based ceramics possess the highest strain of ~0.7% at 50 kV/cm, but this is accompanied by large hysteresis, which is attributed to the transition between the ergodic relaxor and ferroelectric phases, driven by the high electric field (*9*–*12*). Thus, the normalized strain S_{max}/E_{max} (equivalent to the average large signal piezoelectric strain coefficient d_{33}^*) is often below 1000 pm/V, whereas the strain at low electric field is very small in NBT-based ceramics (*9*, *11*). By contrast, (K,Na)NbO_3 (KNN)–based piezoelectric ceramics have a comparable small signal piezoelectric coefficient d_{33} to commercially used Pb(Zr,Ti)O_3 (PZT) ceramics (*6*, *13*). However, current KNN-based ceramics generally need complex stoichiometric control and have electrostrains only up to 0.3%, even when driven by electric fields of 40 to 60 kV/cm (*12*, *14*, *15*). Considering practical actuation applications, the common driving electric field is on the order of 20 kV/cm for commercially used PZT-based ceramics (*16*). Lead-free materials have not exhibited electrostrain properties (including strain value, hysteresis, thermal stability, and fatigue resistance) at driving electric field of 20 kV/cm comparable with those of commercially available PZT materials.

The introduction of defect dipoles, focusing on $B - V_O^{\cdot\cdot}$ (where B refers to the dopant-occupying B-sites of ABO_3 perovskite, such as Mn, Cu, or Fe), was reported to be an effective strategy to enhance electric field–induced strain (*17*–*20*). For example, the defect dipoles in aged $BaTiO_3$ single crystal, where $Fe'_{Ti} - V_O^{\cdot\cdot}$ defect dipoles locally align along the direction of spontaneous polarization, can provide a restoring force to regulate reversible domain switching, resulting in pinned polarization–electric field (*P-E*) hysteresis loops and high strain with large hysteresis (*1*, *18*). In Mn-, Cu-, or Fe-doped KNN ceramics, defect dipoles cause asymmetric strain–electric field (*S-E*) curves (*17*, *19*). Although high bipolar strain of 0.4 to 0.5% (driven by 50 kV/cm) and unipolar strain of 0.18% (driven by 35 kV/cm) have been achieved in Cu- or Fe-doped KNN ceramics, Ar atmosphere–assisted sintering and poling-aging procedure are essential (*19*). Thus, lead-free ceramics with simple compositions, low cost, facile processing, and high electrostrain performance are highly desired.

We introduce $V'_{K/Na} - V_O^{\cdot\cdot}$ defect dipoles to achieve excellent strain performance in the simple composition $[K_{0.5(1-x)}Na_{0.5(1-x)}Sr_x]NbO_3$ (KNSN100x) by means of the solid-state reaction method without any special treatments such as poling-aging (*21*). Giant bipolar electrostrains of 1.05% at room-temperature and 1.67% at 160°C driven by 50 kV/cm are obtained in KNSN3 (x = 0.03) ceramics, which are very high values for lead-free ceramics and comparable to those of lead-free or lead-based single crystals (*1*, *22*, *23*). For practical application scenarios, KNSN3 ceramics also possess higher unipolar strain (0.25% @ 20 kV/cm) than any other lead-free ceramics or commercially available PZT-based ceramics. Of particular interest is that KNSN3 ceramics exhibit good fatigue resistance, low hysteresis, and thermal stability, showing great potential to replace lead-based piezoelectric ceramics for actuator applications. The microstructural and electrical characterizations show that their excellent strain performance originates from a one-way–trip alignment of $V'_{K/Na} - V_O^{\cdot\cdot}$ defect dipoles driven by external electric field and the interaction between defect dipoles and domain switching.

We used x-ray diffraction (XRD) patterns for KNSN ceramics to investigate the composition dependence of the crystal structure (Fig. 1A), where the crystallographic indexing refers to the pseudocubic unit cell. Clearly, all KNSN ceramics have a pure perovskite structure, implying that the Sr ions have diffused into the KNN lattice. The enlarged view of {200} Bragg peaks indicates that the Sr doping has a substantial effect on average symmetry. According to the peak profiles, KNSN2 mainly presents an orthorhombic phase, whereas KNSN6 mainly presents a cubic-like phase. The Rietveld refinement results (fig. S1 and table S1) indicate multiphase coexistence in KNSN ceramics. As Sr content increases, the fraction of orthorhombic phase decreases, whereas the cubic-like phase increases; additionally, the tetragonal phase is found in KNSN3, KNSN4, and KNSN6 ceramics, indicating that Sr can enhance average lattice symmetry. The XRD results agree with the temperature-dependent dielectric measurements (Fig. 1B), in which the peaks of characteristic temperatures for the orthorhombic-etragonal transition (T_{O-T}) and the tetragonal-cubic transition (T_C) are relatively sharp in KNSN2 ceramics, broaden substantially in KNSN3 and KNSN4, and nearly disappear in KNSN6 ceramics.

We measured *P-E* hysteresis loops and corresponding current–electric field (*I-E*) curves for KNSN samples (x = 0.02 to 0.06, where x represents Sr content) (Fig. 1C and fig. S2A).

[1]State Key Laboratory of Metal Matrix Composites, School of Materials Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China. [2]State Key Laboratory of High Performance Ceramics and Superfine Microstructures, Shanghai Institute of Ceramics, Chinese Academy of Sciences, Shanghai 200050, China. [3]School of Physical Science and Technology, Shanghai Tech University, Shanghai 201210, China. [4]Institute for Superconducting and Electronic Materials, Australian Institute for Innovative Materials, University of Wollongong, Wollongong, NSW 2500, Australia. [5]Artificial Crystal Research Center, Shanghai Institute of Ceramics, Chinese Academy of Sciences, Shanghai 201800, China. [6]Department of Materials Science and Engineering, Virginia Tech, Blacksburg, VA 24061, USA.
***Corresponding author. Email: fmail600@mail.sic.ac.cn (Zhengqian Fu); shujun@uow.edu.au (Shujun Zhang); ypguo@sjtu.edu.cn (Yiping Guo)**
†These authors contributed equally to this work.

Higher Sr content (x = 0.04, 0.06) leads to high lattice symmetry and very weak ferroelectricity, whereas the KNSN2 and KNSN3 can achieve robust ferroelectricity. Adding Sr is beneficial to obtaining high-density KNSN ceramics (figs. S2, B and D) and is conductive to an enhanced breakdown strength. KNSN3 samples have the highest small-signal piezoelectric charge coefficient d_{33} (fig. S2B) because of strong ferroelectricity with multiphase coexistence. KNSN samples (x = 0.02, 0.03, 0.04) exhibit asymmetric bipolar *S-E* curves that are similar to the $B - V_O^{\cdot\cdot}$ defect dipoles–modified KNN ceramics (*19*), whereas the strain of KNSN6 samples show a conventional symmetric *S-E* curve (Fig. 1D). Additionally, KNSN3 ceramics exhibit a giant bipolar strain of 1.05% at 50 kV/cm (fig. S3), corresponding to a piezoelectric strain coefficient of 2100 pm/V, which is a high value for a lead-free piezoceramic. The detailed electrostrain performances (including test conditions, strain values, and corresponding piezoelectric strain coefficients S_{max}/E_{max}) are summarized in table S3. Next, we will demonstrate that the giant asymmetric strain observed in KNSN3 is caused by the synergistic effect of the aligned stable $V'_{K/Na} - V_O^{\cdot\cdot}$ defect dipoles, the multiphase coexistence and strong ferroelectricity.

Point defects of K^+/Na^+ vacancies ($V'_{K/Na}$) and oxygen vacancies ($V_O^{\cdot\cdot}$) are generally believed to be unavoidable in KNN-based ceramics. We used electron paramagnetic resonance spectroscopy (fig. S2C) to confirm the existence of $V_O^{\cdot\cdot}$ in KNSN ceramics and we found that Sr can reduce the concentration of $V_O^{\cdot\cdot}$ as a donor dopant ($Sr^{\cdot}_{K/Na}$). Chemical composition test results (table S2) also imply that the concentration of $V_O^{\cdot\cdot}$ decreases with increasing Sr content (x = 0.01 to 0.06; the content of $V_O^{\cdot\cdot}$ in pure KNN is slightly lower than that of KNSN1 because of its lower sintering temperature). We found that in KNSN ceramics with low Sr content (x = 0, 0.01), the abundant $V_O^{\cdot\cdot}$ and low bulk density (table S2 and fig. S2B) cause substantial leakage current and low breakdown strength, making property measurement impossible.

As there are no B_{Nb} defects introduced in KNSN, the asymmetric *S-E* curves are considered to be induced by $V'_{K/Na} - V_O^{\cdot\cdot}$ or $V'_{K/Na} - Sr^{\cdot}_{K/Na}$ defect dipoles, in which K^+/Na^+ vacancies ($V'_{K/Na}$) exist because of the volatilization of alkali metal oxides during high- temperature sintering (*24*). Analogous to $B_{Nb} - V_O^{\cdot\cdot}$ defect dipoles (*17*, *19*, *20*), $V'_{K/Na} - V_O^{\cdot\cdot}$ or $V'_{K/Na} - Sr^{\cdot}_{K/Na}$ is also expected to affect the strain behavior. The energy assumption of $V'_{K/Na} - Sr^{\cdot}_{K/Na}$ reorientation (2.34 eV) is much higher than that of $V'_{K/Na} - V_O^{\cdot\cdot}$ (0.90 eV) (fig. S4). Meanwhile, the $V'_{K/Na} - Sr^{\cdot}_{K/Na}$ defect dipoles are along <001> directions, deviating from the easy polarization rotation path in the orthorhombic phase, which possesses <110> spontaneous polarization directions (*18*). Therefore, we assume that $V'_{K/Na} - V_O^{\cdot\cdot}$ rather than $V'_{K/Na} - Sr^{\cdot}_{K/Na}$ defect dipoles play a major role in KNSN ceramics.

The differences in *S-E* behavior between KNSN3 and PZT are clear when plotted (fig. S5). The electric field–induced strain in ferroelectric materials is closely related to polarization with quadratic dependence (*12*), and the non-180° domain switching causes typical butterfly-shaped *S-E* curves (fig. S5, A and D). However, KNSN3 samples show an *S-E* curve that closely resembles their *P-E* hysteresis loop under 20 kV/cm (fig. S5, E and F). We found three distinctive features compared with those of PZT ceramics: first, KNSN3 samples have two strain values at zero field that correspond to two different strain states (*a* and *d*) (fig. S5F). Second, the two states can be switched by external electric field when synchronized with the polarization. As we observed (fig. S5, E to G), the strain state switches between *a* and *d* states if polarization switch occurs, and in this case, the samples exhibit higher strain with larger hysteresis. Third, the strain response of KNSN3 sample to unipolar electric fields with opposite directions are also contrary, in which negative strain is achieved under positive applied electric field, and vice versa.

To understand the defect dipole contribution to the giant electrostrain, we investigated the polarization and strain behaviors of KNSN3 samples in detail, with the first measurements presented all from fresh samples (Fig. 2). After the first application of unipolar electric field, the KNSN3 samples show a remnant polarization (P_r) of 25 μC/cm^2 (Fig. 2A). However, the P_r ends at –3 μC/cm^2 instead of at –25 μC/cm^2 under the reversed electric field (Fig. 2A, second run). In the third run, the P_r returns to 25 μC/cm^2 under positive electric field. The three test results indicate that the overall P_r during the first run contains a fixed part of ~11 μC/cm^2, which is comparable with the reversible part of ~14 μC/cm^2, the latter originating from the ferroelectric domain switching. This phenomenon is not observed in normal ferroelectric materials (take PZT-5H as an example) (fig. S6). The piezoresponse force microscopy (PFM) characterization further demonstrates that the polarization can indeed be nearly reversed by the electric field (fig. S7). Thus, the fixed part should result from the aligned $V'_{K/Na} - V_O^{\cdot\cdot}$ defect dipoles along the external field during the first run, which is subsequently stabilized. In the ferroelectric lattice, defect dipoles exhibit high polarizability, several orders of magnitude higher than ferroelectric polarization (*19*), so they can potentially produce such a high amount of fixed polarization when they are aligned. According to the direction of the electric field, the domain switching occurred only in the first, third, and fifth runs, whereas there was almost no domain switching involved in the second and fourth runs (Fig. 2B). In the first and

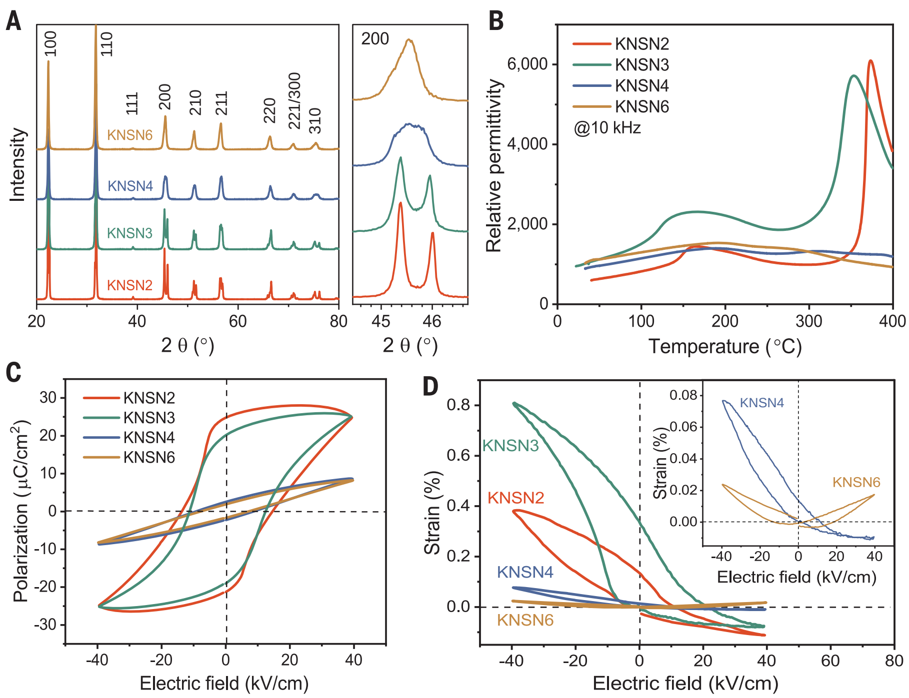


Fig. 1. Composition-dependent crystal structure and electric properties of KNSN ceramics. (**A**) X-ray diffraction (XRD) patterns. (**B**) Temperature dependence of the relative permittivity over temperature range of 25° to 400°C measured at 10 kHz. (**C**) Polarization–electric field (*P-E*) hysteresis loops. (**D**) Strain–electric field (*S-E*) curves of KNSN ceramics at 40 kV/cm. The inset is the enlarged parts of KNSN4 and KNSN6.

second runs, the strain behavior was consistent with that of a normal ferroelectric, and possesses the following two features: the strain starts to increase only after the electric field exceeds coercive field (about –9.9 kV/cm). The first *S-E* curve has large hysteresis, whereas the subsequent second one shows much smaller hysteresis. However, typical KNN-based ceramics with domain switching alone cannot contribute to such a giant strain value of 0.58% at –20 kV/cm. Although the domain switching indeed occurs, the strain does not present the well-known butterfly shape. In the third run, the strain should increase for domain switching alone but continues to decrease after coercive field, and in the fifth run, the strain should decrease for domain switching alone but continues to increase before coercive field.

We conclude that the interaction between aligned defect dipoles (P_d) and domain switching should be responsible for the strain behavior and giant electrostrain we observed (Fig. 2, A and B). We propose an explanation that includes simply superimposing the lattice stretching or contraction from P_d with the butterfly curve from domain switching (fig. S8). At the *a* point, both defect dipoles and polarization are aligned after the initial electric field stimulus. By applying positive electric field in an *a-b-c* route, the P_d causes continuous lattice contraction in both *a-b* and *b-c* routes, whereas the domain switching leads to strain decrease in the *a-b* route but strain increase in the *b-c* route. Thus, the overall strain exhibits an inflection point (point *b*) at the coercive field of the *a-b-c* route—a rapidly descending *a-b* segment and a slowly descending *b-c* segment. With the removal of the external field in the *c-d* route, the defect dipole–induced lattice contraction gradually weakens and leads to an increase in strain, whereas the aligned domains partially recover and lead to a decrease in strain. Thus, the overall strain stays nearly unchanged over the *c-d* route. When the strain curve goes from point *a* to *d*, the 180° polarization switch occurs while the P_d remains fixed. In this case, with applying the negative electric field, the *d-e* segment shows a slow increase in strain because the P_d leads to lattice stretch but domain switching leads to strain decrease, while the *e-f* segment shows a rapid increase in strain because the P_d continuously gives rise to lattice stretch and domain switching also leads to strain increase. When the electric field is removed over the *f-a* route, the lattice stretching by P_d gradually weakens and leads to a strain decrease, whereas the aligned domains partially recover and cause strain decrease too. Thus, the overall strain shows an obvious decrease during the *f-a* route, which is different from that of the nearly constant *c-d* route. The difference between *f-a* and *c-d* segments reveals that the P_d is fixed; otherwise, the strain curve slopes of the segments should be similar. The substantial strain difference between states *a* and *d* (the two zero-field states in *S-E* curve) originates from the interaction between ferroelectric domains and P_d. That is, defect dipoles generate additional lattice stretching or contraction under external field beyond ferroelectricity.

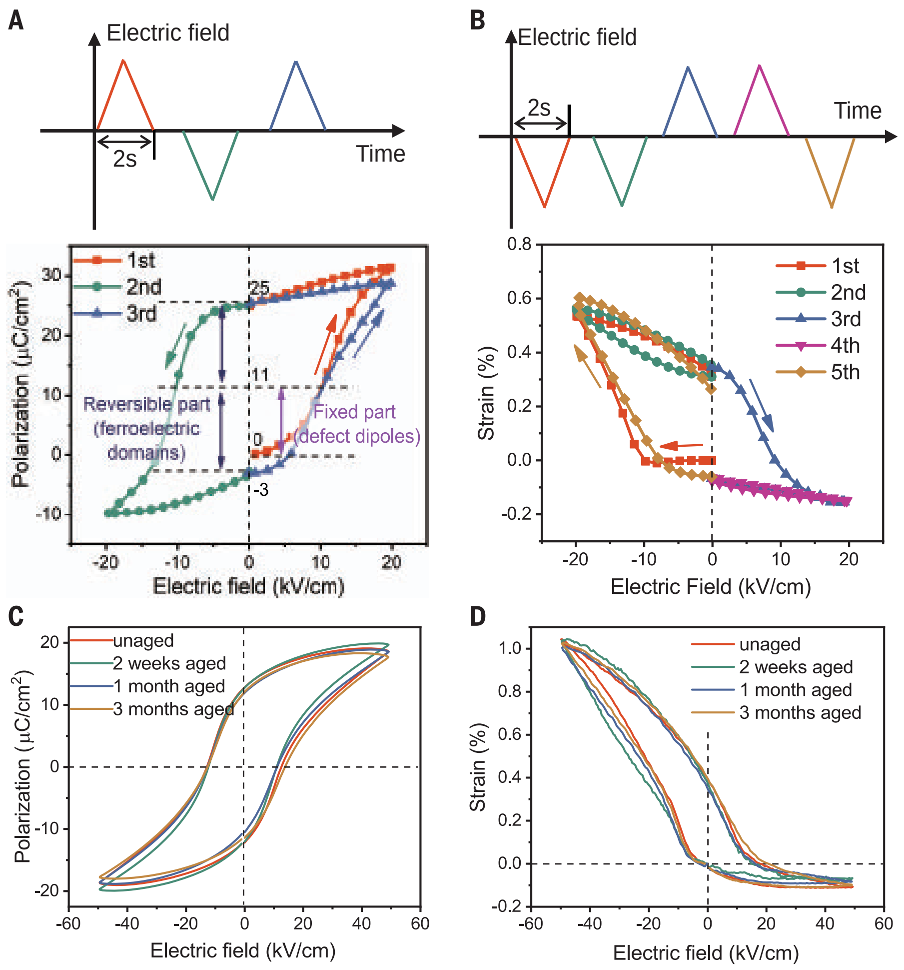


Fig. 2. The effects of defect dipoles on polarization and strain behaviors in KNSN3. (**A** and **B**) *P-E* behavior and *S-E* behavior of the initial few unipolar electric field applications on a fresh sample with corresponding wave profiles, respectively. The unipolar measurements were carried out in 2 seconds to clearly show the *P-E* behavior, so the waveform can be regarded as part of an alternating triangle wave with a frequency of 0.25 Hz. (**C** and **D**) *P-E* hysteresis loops and *S-E* curves during a 3-month aging process, respectively.

To further examine the interaction model between P_d and domain switching, we designed a poling-aging experiment (fig. S9). The fresh samples were first poled to align the defect dipoles, then poled by the reversal electric field at 120°C, followed by 14 days of aging. If the time of reversal poling is short (fig. S9A), or if the poling field is not high enough (fig. S9B), the domains reverse, and the defect dipoles do not. Thus, after aging, the strain behavior is consistent with the fifth run in Fig. 2B (domain switching and lattice stretching induced by defect dipoles). The defect dipoles can only be forced to reverse by prolonged high reversal fields poling, and aging (fig. S9C), in which the strain behavior is consistent with third run in Fig. 2B after these harsh treatments. The results shown in fig. S9 further verify the interaction between domains and defect dipoles.

After the P_d is formed by external electric field, its high stability is evidenced by the poling-aging results and the direct-current strain behavior. After 3 months of aging, the *P-E* and *S-E* loops have minimal change (Fig. 2, C and D). The P_d are hard to reorient (fig. S9C), thus verifying the long-term stability of the P_d. Additionally, the aging process will not enhance the electrostrain value, indicating that all the defect-dipole alignments are achieved in the first application of electric field (Fig. 2D). Furthermore, the P_d cannot be reoriented by low-frequency stimuli even in the direct current condition (fig. S5G).

To understand the effect of $V'_{K/Na} - V_O^{\cdot\cdot}$ defect dipoles on strain behavior, we further

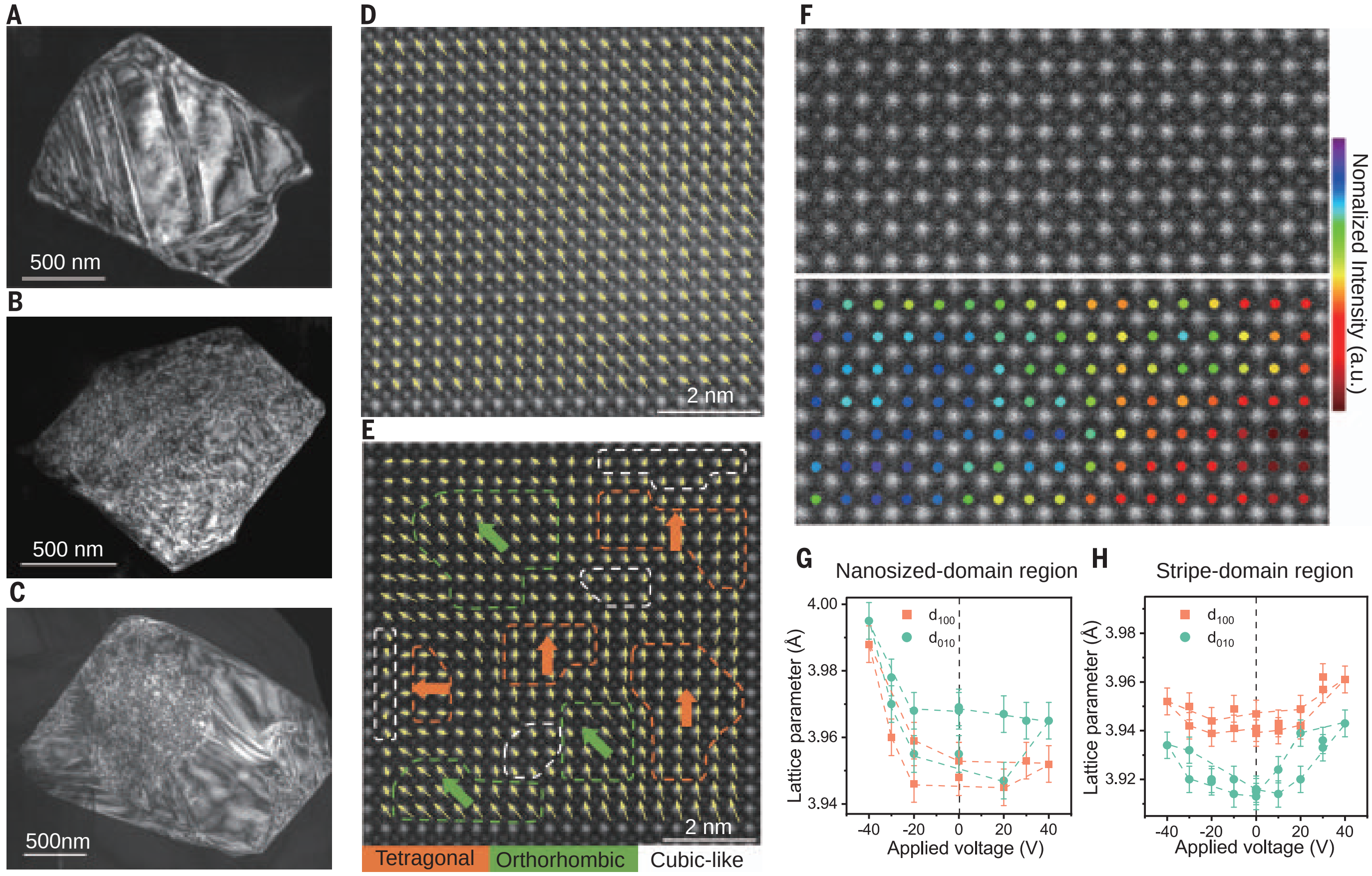


Fig. 3. Microstructure characterizations of KNSN3. (**A** to **C**) The TEM dark-field images for stripe, nanosized, and mixed domains, respectively. (**D** and **E**) Atomic-scale HAADF images of a stripe domain region and a nanosized domain region, respectively, where horizontal and vertical directions correspond to two perpendicular <100> crystallographic directions. The mixing of various atomic displacement states suggests multiphase coexistence in nanosized domain region. (**F**) The original atomic-scale HAADF image and the image marked with normalized A-site intensity by color of the same area. (**G** and **H**) Evolution of interplanar spacing of {100} planes during in situ electric stimuli of a stripe domain region and a nanosized domain region, respectively. The error bars represent the standard deviations of five times the measurements from the experimental images.

investigated the microstructure of KNSN3 ceramics. Representative domain structures for KNSN3 ceramics, which exhibit three different morphological features, are shown in Fig. 3, A, B, and C. The first feature is stripe domains (Fig. 3A), of which the domain walls mainly lie in {110} or {001} planes. This crystallographic characteristic indicates that the areas with stripe domains are consistent with orthorhombic symmetry. The second feature is nanosized domains (Fig. 3B), of which the domain walls do not lie in specific crystal planes, implying a corresponding complex phase structure. The third feature is mixed domains, of which the nanosized domains are usually surrounded by stripe domains (Fig. 3C) which dominate the domain morphological feature in most of the grains. The coexistence of the stripe and nanosized domains is also found in KNSN2 (fig. S10), which has asymmetric *S-E* curve (Fig. 1D). The energy dispersive spectrum (EDS) mappings of areas of KNSN3 ceramics containing dozens or singular grains demonstrate an inhomogeneous distribution of K and Na elements (fig. S11), which is a common phenomenon in KNN-based ceramics. The Na-rich regions have additional ½(*ooe*) superlattice reflections (where *o* and *e* denote odd and even Miller indices, respectively) in selected-area electron diffraction (SAED), whereas the K-rich regions only exhibit the main reflections of perovskite structure (fig. S12). According to the previous structural studies on KNN ceramics, the ½(*ooe*) superlattice reflections are most likely the result of the in-phase tilting of oxygen octahedra (*25*).

With the help of spherical aberration-corrected scanning transmission electron microscopy (C_s-STEM), we revealed the structures on the atomic scale (Fig. 3, D and E). Analysis of the local atomic displacement in the high-angle annular dark field (HAADF) image shows that Nb cations are homogenously displaced along the <110> direction in the stripe domain region, which is consistent with the polarization direction of the orthorhombic phase (Fig. 3D). By contrast, both the direction and magnitude of Nb cation displacement vary substantially in nanosized domain regions. According to the projection of the pseudocubic unit cell, the nanosized domain region suggests a multiphase state involving orthorhombic, tetragonal, and cubic-like phases (Fig. 3E), which is consistent with the average XRD results (table S1). We observed a high concentration of $V'_{K/Na}$, which was evidenced by the drastically variable intensity of A-site atomic columns (Fig. 3F), providing a basis for the formation of abundant defect dipoles.

In situ–biased TEM results on nanosized domain regions and stripe domain regions indicate the microscopic evolution of their lattices under electric field. Figure 3G shows the evolution of measured lattice parameters along two orthogonal <100> directions in the pure stripe domain grain, where the lattice exhibits symmetric change under the electric fields along both directions, which is consistent with

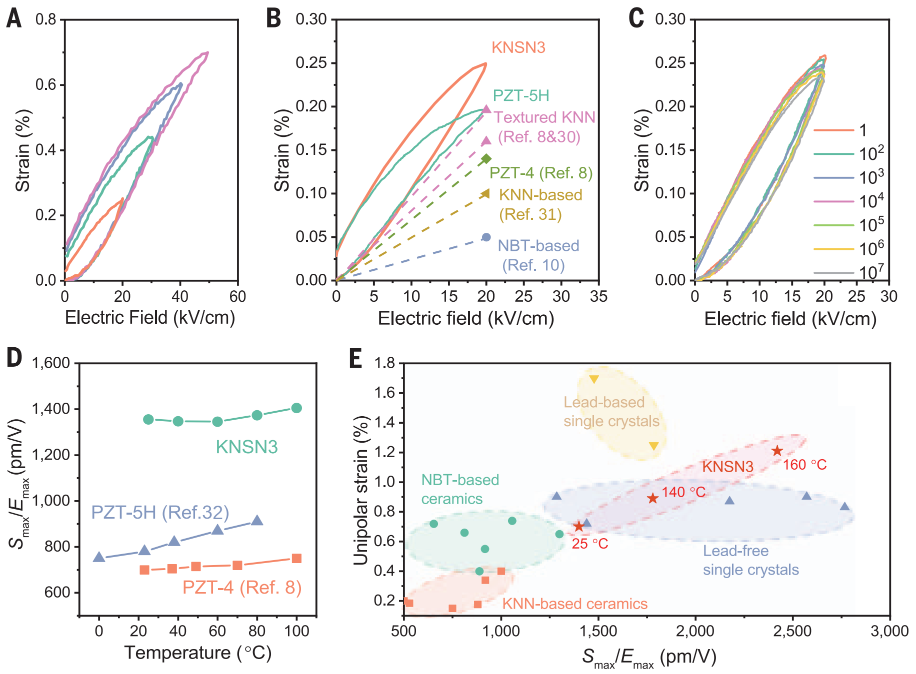


Fig. 4. Excellent electrostrain properties of KNSN3. (**A**) Unipolar strains of KNSN3 at electric field from 20 kV/cm to 50 kV/cm. (**B**) Comparison of unipolar strain at 20 kV/cm among piezoceramics (*8*, *10*, *30*, *31*). (**C**) Unipolar 20 kV/cm fatigue at room temperature. See table S4 for details of fatigue and test conditions. (**D**) Thermal stability comparison between KNSN3 and commercial lead-based ceramics (*8*, *32*). (**E**) Unipolar Strain performance comparison among representative piezoelectrics with giant electric-field-induced strains (*1*, *2*, *7*, *9*–*11*, *19*, *23*, *33*–*38*).

the strain behavior observed in ferroelectric domains. By contrast, the changes in local lattice parameters of the pure nanosized domain grain (Fig. 3H) reveal that the lattice parameters decrease under positive electric stimuli, while they clearly increase under negative electric stimuli. The stretching and contraction of lattices in nanosized domain regions are in good agreement with the shape of *S-E* curves (Fig. 1C).

According to the Landau-Ginzburg-Devonshire phenomenological theory, the free energy profile becomes flat when different ferroelectric phases coexist, representing a low barrier for polarization rotation among different states (*26*). Because of the multiphase coexistence in nanosized domain regions, rotation and alignment of the defect dipoles along the electric field during the first application of the external field is energetically favorable. According to the theory proposed by Arlt and Neumann (*27*), P_d will introduce an extra electric field on the free energy profile (fig. S13), causing the system to descend into a potential well in which the ferroelectric polarization of the nanosized domain regions align along the same direction as P_d. This observation is also supported by the "collinear" polar nanoregions where the electrostatic, gradient, and elastic energies keep them stable after reorientating along each other (*28*). Therefore, the defect dipoles in nanosized domain regions will remain fixed after the alignment, which results in the fixed part of the observed polarization (Fig. 2A). In addition, compared with Cu- or Fe-doped KNN ceramics (*19*), the $V'_{K/Na} - V_O^{\cdot\cdot}$ defect dipoles have the same direction with the spontaneous polarization in the orthorhombic phase of KNSN ceramics. These characteristics allow the $V'_{K/Na} - V_O^{\cdot\cdot}$ defect dipoles to require no aging for alignment in multiphase coexistence with nanosized domains. When an external electric field is applied, the defect dipoles drive stretching or contraction of the surrounding lattice and interact with the domain switching, leading to the asymmetric *S-E* curve, even under high electric field conditions. More recently, similar strain behavior has also been reported in NBT ceramics (*29*), in which ultrahigh strain is associated with switching between disordered and ordered polarization states, complex structure with multiphase coexistence, non-180° switching, and domain stabilization. The polarization in KNSN3 samples, however, can be almost reversed by the electric field (fig. S7). Thus, we believe that the coupling of stabilized $V'_{K/Na} - V_O^{\cdot\cdot}$ defect dipoles aligning along <110> direction and ferroelectric domain switching is responsible for the observed high asymmetric strain level.

For actuator applications, electrostrain is usually evaluated from the unipolar strain value. We achieved a unipolar strain as high as 0.7% at 50 kV/cm (Fig. 4A) (detailed data are summarized in table S3). Considering the common driving electric field of 20 kV/cm, the unipolar S-E curves and strain values for typical lead-based or lead-free piezoceramics are compared in Fig. 4B. KNSN3 ceramics clearly have the highest strain value (0.25%). Unlike the large-strain hysteresis observed in aged $BaTiO_3$ single crystal (reversible domain switching) and NBT-based ceramics (field-induced phase transition) (*10*, *18*), low hysteresis is observed when the external field, polarization, and defect dipoles are all in the same direction (Fig. 2B, second run), with the domain switching nearly not involved. Therefore, such a unipolar strain with low hysteresis (31% at 20 kV, comparable with commercially used PZT ceramics) is expected to benefit piezoelectric actuation. In addition, the high stability of the P_d will benefit fatigue resisance. We characterized the cycling reliability of strain under 20 kV/cm unipolar electric field, at which the unipolar strain only degrades 9% after 10^7 cycles (Fig. 4C) (the detailed data of fatigue tests are summarized in table S4). KNSN3 ceramics have excellent thermal stability (Fig. 4D), with unipolar strain variation of less than 5% over a temperature range of 25°C to 100°C, comparable with PZT-4 and better than PZT-5H ceramics. Therefore, KNSN3 ceramics have immense potential to replace Pb-based piezoelectric ceramics, considering their capability for giant strain, low driving electric field, good fatigue resistance, and thermal stability.

When the temperature exceeds 100°C, the electric field–induced strain of KNSN3 ceramics further increases with the increase in temperature (fig. S14). The highest electrostrains—1.67% for bipolar and 1.2% for unipolar, both very high values for piezoceramics (table S3)—were achieved at 160°C. We also observed the good fatigue resistance of KNSN3 at 160°C (fig. S14C and table S4). These performances prove the great potential of KNSN3 ceramics for high-temperature actuator applications, such as piezoelectric injectors in diesel engines (*1*). We also considered whether polymorphic phase transition is the reason for the observed enhanced strain (fig. S14, D and E). The diffused dielectric anomaly that occurs around 160°C suggests the orthorhombic-tetragonal phase transition, whereas the disappearance of the Raman mode at 200 cm^{-1} also confirms the change in crystal symmetry (*2*). We expect even higher strain at room temperature may be obtained by decreasing $T_{O\text{-}T}$—for example, by means of Li doping (*14*).

We provide a comparison of unipolar strain performance of the studied KNSN3 ceramics with representative piezoelectrics (Fig. 4E), in which KNSN3 is superior to the lead-free ceramics and even comparable with those of

both lead-free and lead-based single crystals. The underlying mechanism responsible for the giant electrostrain is the coupling of the defect dipoles with ferroelectric domains, which is done by tailoring the $V'_{K/Na} - V_O^{\cdot\cdot}$ defect dipoles and microstructure, thus providing a paradigm for the design of giant-strain piezoelectric materials. Considering the high strain under 20 kV/cm, good fatigue resistance, and thermal stability, KNSN is expected to be a great potential lead-free alternative for broad-temperature range and high-displacement piezoelectric actuator applications.

REFERENCES AND NOTES

1. P. Fan *et al*., *J. Materiomics* **7**, 508–544 (2021).
2. J. Hao, W. Li, J. Zhai, H. Chen, *Mater. Sci. Eng. Rep.* **135**, 1–57 (2019).
3. "Piezoelectric devices market with COVID-19 impact analysis by material (piezoelectric ceramics, polymers), product (piezoelectric actuators, transducers, motors), application (aerospace & defense, industrial, consumer), and region-global forecast to 2026," *Piezoelectric Devices Market* (MarketsandMarkets, 2021).
4. "Lead-free piezoelectric ceramics: Technologies and global opportunities," (BCC Research, 2019).
5. B. Narayan *et al*., *Nat. Mater.* **17**, 427–431 (2018).
6. T. Zheng, J. Wu, D. Xiao, J. Zhu, *Prog. Mater. Sci.* **98**, 552–624 (2018).
7. Y. Wang *et al*., *Adv. Electron. Mater.* **6**, 1900949 (2020).
8. Y. Saito *et al*., *Nature* **432**, 84–87 (2004).
9. T. Li *et al*., *Acta Mater.* **128**, 337–344 (2017).
10. X. Liu, X. Tan, *Adv. Mater.* **28**, 574–578 (2016).
11. J. Wu *et al*., *Nano Energy* **76**, 105037 (2020).
12. W. Jo *et al*., *J. Electroceram.* **29**, 71–93 (2012).
13. Q. Liu *et al*., *Energy Environ. Sci.* **11**, 3531–3539 (2018).
14. Y. Guo, K.-i. Kakimoto, H. Ohsato, *Appl. Phys. Lett.* **85**, 4121–4123 (2004).
15. M. Waqar, H. Wu, J. Chen, K. Yao, J. Wang, *Adv. Mater.* **34**, 2106845 (2021).
16. C. D. Near, paper presented at the 1996 Symposium on Smart Structures and Materials, San Diego, CA, United States, 1 May, 1996.
17. Z.-H. Zhao, Y. Dai, F. Huang, *Sustainable Mater. Technol.* **20**, e00092 (2019).
18. X. Ren, *Nat. Mater.* **3**, 91–94 (2004).
19. Z. Zhao, Y. Lv, Y. Dai, S. Zhang, *Acta Mater.* **200**, 35–41 (2020).
20. J. Yin, C. Li, B. Wu, Z. Li, J. Wu, *J. Eur. Ceram. Soc.* **41**, 2506–2513 (2021).
21. Materials and methods are available as Supplementary Materials on *Science* Online.
22. S.-E. Park, S. Wada, L. E. Cross, T. R. Shrout, *J. Appl. Phys.* **86**, 2746–2750 (1999).
23. H. Zhang *et al*., *J. Mater. Chem. C Mater. Opt. Electron. Devices* **2**, 10124–10128 (2014).
24. Y.-X. Liu, H.-C. Thong, Y.-Y.-S. Cheng, J.-W. Li, K. Wang, *J. Appl. Phys.* **129**, 024102 (2021).
25. D. W. Baker, P. A. Thomas, N. Zhang, A. M. Glazer, *Acta Crystallogr. B* **65**, 22–28 (2009).
26. F. Li *et al*., *Nat. Mater.* **17**, 349–354 (2018).
27. G. Arlt, H. Neumann, *Ferroelectrics* **87**, 109–120 (1988).
28. F. Li *et al*., *Nat. Commun.* **7**, 13807 (2016).
29. W. Feng *et al*., *Nat. Commun.* **13**, 5086 (2022).
30. P. Li *et al*., *Adv. Mater.* **30**, 1705171 (2018).
31. X. Lv, J. Wu, *J. Mater. Chem. C Mater. Opt. Electron. Devices* **7**, 2037–2048 (2019).
32. D. Wang, Y. Fotinich, G. P. Carman, *J. Appl. Phys.* **83**, 5342–5350 (1998).
33. S.-Y. Choi *et al*., *Chem. Mater.* **24**, 3363–3369 (2012).
34. Q. Wei, M. Zhu, M. Zheng, Y. Hou, *J. Alloys Compd.* **782**, 611–618 (2019).
35. L. Zhang *et al*., *Adv. Funct. Mater.* **30**, 2004641 (2020).
36. G. Dong *et al*., *J. Materiomics* **7**, 593–602 (2021).
37. H. Zhang *et al*., *Scr. Mater.* **75**, 50–53 (2014).
38. Z. Dong *et al*., *J. Am. Ceram. Soc.* **105**, 6744–6754 (2022).

ACKNOWLEDGMENTS

F. Wang of Shanghai Normal University is acknowledged for the help in PFM characterizations. The authors thank T. Silver for language editing. **Funding:** Natural Science Foundation of China (No. 52032012), Basic Research Project of Science and Technology of Shanghai (No. 20JC1415000). **Author contributions:** Y.G. came up with the original idea; G.H., F.Z., S.Z. and Y.G. directed the project; G.H., K.Z. and B.W. performed the experiments; G.H., B.W. and J.W. prepared the ceramic materials used in this study; G.H., K.Z., Z.F. and Y.G. wrote the original draft of the manuscript; F.X., H.L., D.V. and S.Z. supervised this work and revised the manuscript. **Competing interests:** Authors declare that they have no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials.

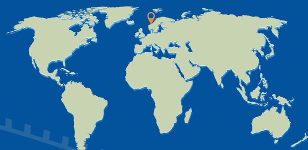
License information: Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.ade2964
Materials and Methods
Figs. S1 to S14
Tables S1 to S4
References (*39–46*)

Submitted 6 August 2022; accepted 8 November 2022
10.1126/science.ade2964

63 north
KNOWLEDGE FOR A BETTER WORLD



THE ONSAGER FELLOWSHIPS

16 tenure-track positions available at NTNU

The Norwegian University of Science and Technology (NTNU) is Norway's primary institution for educating the future's engineers and scientists. The university also has strong programmes in the social sciences, teacher education, the arts and humanitites, medicine, architecture and fine art. NTNU's cross-disciplinary research delivers creative innovations that have far-reaching social and economic impact and that help contribute to a better world.

The Onsager Fellowship programme at NTNU is designed to attract the most talented scholars with an established reputation for high quality research and a commitment to learning and teaching at the university level.

NTNU | Norwegian University of Science and Technology

APPLY FOR A TENURE-TRACK POSITION AS AN ASSOCIATE PROFESSOR IN:

- Quantum chemistry
- Computational Porous Media Physics
- Socially-innovative cities
- Circular Building Design
- Experimental Environmental Psychology
- Political Behaviour
- Solid Mechanics
- Marine structures
- European Studies
- STS and sustainability transitions
- Post-quantum Cryptography
- Experimental Quantum Communication
- Artificial Intelligence (AI) and Health
- Brain MRS
- Public and Political Economics
- Biosystematics

More info at: www.ntnu.edu/onsager-fellowship/
Closing date: 1 February 2023

TOP RESEARCH DEMANDS BRILLIANT MINDS – WE'RE ALWAYS LOOKING FOR THE BEST

GLENN FOUNDATION FOR MEDICAL RESEARCH | afar american federation for aging research

2023 FUNDING OPPORTUNITIES

RESEARCH GRANTS FOR JUNIOR FACULTY
$150K, 1-2 years / LOI deadline:12.15.22

POSTDOCTORAL FELLOWSHIPS IN AGING RESEARCH
$75k, 1 year / LOI deadline: 01.25.23

GLENN FOUNDATION DISCOVERY AWARDS
$525K, 3 years / LOI deadline: 02.15.23

For Research Focus, Eligibility Requirements, and More, visit **www.afar.org/funding-opportunities**

Northeastern University

TENURED/TENURE-TRACK ASSOCIATE AND ASSISTANT PROFESSOR POSITIONS

The Center for Drug Discovery (CDD) at Northeastern University (NU) invites applications for **two junior (pre-tenure Assistant Professor) and two senior (Tenure-on-Entry Associate or Full Professor) positions with projected start dates in 2023.** The mission of the CDD is to engage in drug discovery and development that will lead to new medications, particularly involving brain disorders, infectious diseases, and cardiovascular disorders. To accomplish this, the CDD is directed to develop novel methodologies to improve drug discovery. Junior scientists using breakthrough technologies are welcome to apply as are senior faculty with internationally recognized programs and strong funding, whose expertise will complement activities of faculty affiliated with the CDD toward drug discovery.

A goal of this recruitment effort is to establish CDD as an NIH-funded Center of Excellence relevant to the mission of institutes that represent the Center's membership. The CDD currently holds a NIDA T32 training grant for graduate students and postdoctoral fellows available to trainees of CDD faculty. The CDD also holds a NIDA R13 grant that sponsors the annual Chemistry and Pharmacology of Drug Abuse Conference at NU. The administrative team of the CDD provides pre- and post-award support to its members. The impact of the CDD is dependent on the outstanding quality of its membership and the strong interactions they develop with each other. Each senior CDD member is a founder of an active start-up company.

Interested candidates should apply with a cover letter, curriculum vita that includes a list of publications, statements addressing Research, Teaching and Equity, and names & contact information for at least three professional references. For more detailed information on these positions and how to apply please consult the University posting:

https://northeastern.wd1.myworkdayjobs.com/careers/job/Boston-MA-Main-Campus/Assistant-Professor--Associate-Professor--Full-Professor_R110627

Successful candidates will have the opportunity to select among four participating Colleges and multiple departments for their academic home. Joint appointments in departments across different Colleges are highly encouraged at NU. NU is an equal opportunity employer, seeking to recruit and support a broadly diverse community of faculty and staff.

What's Your Next Career Move?

From networking to mentoring to evaluating your skills, find answers to your career questions on *Science* Careers

To view the complete collection, visit
ScienceCareers.org/booklets



ScienceCareers

FROM THE JOURNAL SCIENCE AAAS



TULANE UNIVERSITY
SCHOOL of SCIENCE & ENGINEERING

The Department of Ecology & Evolutionary Biology (EEB) at Tulane University invites applications for a **faculty position in Ichthyology (associate or full professor rank**) who will also serve as Director of the Tulane University Biodiversity Research Institute (TUBRI) and Curator of the Royal D. Suttkus Fish Collection.

See **https://sse.tulane.edu/eebio/about/positions/Professor-of-Ichthyology** for additional details about the position, the department and TUBRI.

To apply please visit **https://apply.interfolio.com/117771**. Review of applications will begin **January 2, 2023**. The search will remain open until the position is filled.

Tulane University is an Af rmative Action/Equal Employment Opportunity/ADA Employer. Women and minorities are encouraged to apply.



Features in myIDP include:

- Exercises to help you examine your skills, interests, and values.
- A list of 20 scientific career paths with a prediction of which ones best fit your skills and interests.

myIDP

Visit the website and start planning today!
myIDP.sciencecareers.org

ScienceCareers | AAAS In partnership with:

FASEB Federation of American Societies for Experimental Biology
UMASS. University of Massachusetts Medical School
UCSF University of California San Francisco
UIC
BURROUGHS WELLCOME FUND

By **Madeline Schultz**

Working through grief

As I stepped out of the chemistry building, a chirp from my phone alerted me to a text message from my mom. "Are you still at work? Call me when you get home." My stomach sank as I saw three missed calls: two from my mom and one from my best friend Gabs. After a half-dozen rings, I got my mom's answering machine. The pit in my stomach deepened when I called Gabs, who choked out, "Mads, you have to call your momma." My hands were sweating as I gripped the steering wheel on my short drive home. As soon as I stepped out of my car, my mom called me back. "Daddy had a heart attack while he was surfing today, he's dead, I'm so sorry." I sunk to my knees in the grass and vomited repeatedly, shaking with sobs.

That evening, I booked a one-way ticket to my mom. I emailed my adviser on the way to the airport. His response brought a wave of relief: "I am so sorry! Please take whatever time you need. I hope you are able to get home quickly." The next few weeks were a blur of celebrations of life, lawyers, and meals I could hardly stomach. My research barely crossed my mind, and my adviser reminded me with unconditional kindness, "Please don't feel any pressure to rush back to lab."



"Not being able to tell my dad dulled the accomplishment, but I know he would be proud."

But after a month away, I felt the best thing for me would be to try to force a return to normalcy—which meant returning to school. So, I went back to lab, hoping I could put my grief aside for at least a few hours while I focused on the research I had previously been so enthusiastic about.

But as soon as I returned, I was reminded that my life was still far from normal. As graduate students and professors offered well-meaning sympathy, I felt embarrassed and isolated by my sadness. Seeing the instrument I had been constructing for the past year, I remembered all the times I had FaceTimed my dad, a master craftsman, asking for advice on which tools to use and how to remove a stripped screw without damaging fragile electronics nearby. Instead of getting back to work, I hid in the doorway between two labs to cry.

I felt pathetic: After taking an entire month off, I had expected to tackle my ever-growing to-do list. As the first graduate student in a new group, I felt entirely responsible for keeping our research going. If I didn't step up, I worried our progress would stall.

But despite this self-imposed pressure, about an hour after I arrived in lab that first day I had reached my limit. Sobbing and trembling, I went to the yoga studio. In the heat of the studio and the depth of the stretch, I slowly began to calm.

For the next few weeks, I dragged myself to lab every day, but I was unproductive and uninterested in my research. I began to doubt whether my fascination for science would ever return. Only a year into my Ph.D., I didn't see how I could possibly make it through the years to come. My despair deepened.

Then, a few weeks after returning, a senior graduate student in another lab who was training me on a new instrument asked me how the start of my second year was going. I didn't know her well, but something about her made me feel comfortable opening up. I broke into tears, and she shared with me her experience of losing a loved one during graduate school. It was the first time since I returned that I felt understood.

From then on, I slowly began to feel a bit more normal each day. As my all-consuming grief gradually dulled to a steady hum, my passion for my research returned. While analyzing a new sample or constructing a new piece of my instrument, I began to feel my old excitement again. Three months after returning, I realized I had gone a full week without crying in the space between labs. By piecing together these fleeting moments of productivity, I was able to collect enough data to submit an abstract, which turned into a conference presentation. Not being able to tell my dad dulled the accomplishment, but I know he would be proud.

One year and a few months after his death, I am still learning how to handle my grief, a little more each day. When I look at my instrument now, I feel a deep sadness that I can't share it with my dad. That sadness will never go away. But I also feel proud that I carried on. ■

Madeline Schultz is a Ph.D. student at the University of Rhode Island, Kingston. Send your career story to SciCareerEditor@aaas.org.

ILLUSTRATION: ROBERT NEUBECKER

NEWS FROM Science | AAAS



TRILLIONS OF MICROBES

ONE ESSAY

The NOSTER *Science* Microbiome Prize is an international prize that rewards innovative research by investigators who have completed their terminal degree in the last 10 years and are working on the functional attributes of microbiota. The research can include any organism with the potential to contribute to our understanding of human or veterinary health and disease, or to guide therapeutic interventions. The winner and finalists will be chosen by a committee of independent "scientists" chaired by a senior editor of *Science*. The top prize includes a complimentary membership to AAAS, an online subscription to *Science*, and USD 25,000. Submit your research essay today.



Jennifer Hill, Ph.D.
2022 Winner

NOSTER | Science
MICROBIOME
PRIZE

Apply by 24 January 2023 at www.science.org/noster

Sponsored by Noster Inc